The Dissertation Committee for Min Soon Kim Certifies that this is the approved version of the following dissertation:

Objective Assessment of Aesthetic Outcomes of Breast Cancer Treatment: Quantifying Aesthetic Factors after Breast Reconstruction

Committee:

Mia K. Markey, Supervisor

Alan C. Bovik

Krishnaswamy Ravi-Chandar

Gregory P. Reece

Bugao Xu

Objective Assessment of Aesthetic Outcomes of Breast Cancer Treatment: Quantifying Aesthetic Factors after Breast Reconstruction

by

Min Soon Kim, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2007

Dedicated to

My wife Bosung Jun and son Donovan J. Kim

My parents Sunhong Kim and Wonja Kim

My sisters Kyungjin Kim, Jinah Kim, and Jeongin Kim

My brothers-in law, Jiwoo Park, Kyooseop Kim,

and Kyngduk Kim of blessed memory

Acknowledgements

I would like to give thanks to my supervisor Prof. Mia K. Markey for her sincere advising. Discussions with her helped me to correctly understand the key point of the problems with my research. She has always been supportive of my research from the very beginning point to ending point of my Ph.D. process. Particularly, I truly appreciate Dr. Reece for his help with my understanding of importance of assessment of breast reconstruction. He has been more than a research committee member to me.

Many thanks are due to my dissertation committee: Prof. Alan C. Bovik, Prof. Krishnaswamy Ravi-Chandar, and Prof. Dr. Bugao Xu. Their advice has been of great help for me to complete my doctoral research. I would also like to thank my collaborator Dr. Dr. Karen Basen-Engquist for her advice of subjective scale construction.

I would like to give my hearty thanks to my wife, Bo for her enduring sacrifice and devotion she has made. Particularly, I would like to thank her for changing my life by delivering the most precious gift, my son, Donovan to me. She also changed my life again by reminding me of the existence of the God who always is watching us from a distance.

I would like to give thanks to my parents for their consistent support, encouragement, and prayer. Their love for me is the most important source that enabled me to continue my doctoral study. They have always been beacons in the night. I have

always been proud that I was born of your son and that you are my parents. I would like to thank my parents-in-law for their constant advice and encouragement for years. I will never forget my mother-in-law's prayer for me. The fact that she always prays for me has encouraged me to proceed step by step without giving up.

Also, I would like to thank my dear sisters and brothers-in-law for their wise advice. They have always listened to me very carefully and have given advice that I needed at the moment. I would like to thank my oldest sister for her constant advice and encouragement in spite of all her personal difficulties over the past few years. I would like to say to then that I am so proud of being a brother and uncle of your children.

I also want to say thank you to the BMIL lab members, Jin, Shalini, Mehul, Wendy, Sun, Qiu, and Ernest. I will never forget the time with them. They have been my good friends and peer mentors. The time with my graduate and undergraduate assistants should be appreciated. I want to thank Bill, Haoyu, Nitin, and Justin, for their inspiration and personal help. I also appreciate the technical assistance from our great computer administrators, Al, Zack, and Chris. I thank the BMIL lackeys, Khuyen, Vince, Charles, and Lesley for their Endnote work. I would also like to thank the staff members of BME departments, Joni, Ann, Heidi, Majid, Cheryl, Maggi, and Melanie for their administrative work.

Objective Assessment of Aesthetic Outcomes of Breast Cancer Treatment: Quantifying Aesthetic Factors after Breast Reconstruction

Publication No._____

Min Soon Kim, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Mia K. Markey

Breast cancer is the most common cancer among American women. One in eight women will be diagnosed with breast cancer during her lifetime. Essentially all breast cancer treatment involves surgery. The two most generally performed surgical treatments for breast cancer are breast conservation therapy and mastectomy followed by breast reconstruction. Breast reconstructive surgery is an important component of the breast cancer treatment process.

The aesthetic outcome of breast cancer treatment is a critical factor in breast cancer survivors' quality of life. Aesthetics is a general term that refers to physical characteristics such as symmetry and proportion. Currently, physicians, patients, or other observers evaluate breast aesthetics in a subjective, qualitative manner. However, such assessments are typically based on vaguely defined rating scales that have low intra- and

inter-observer agreement. Their qualitative nature also restricts the analyses that can be performed. Quantitative, objective measures with high reliability are needed to meaningfully relate patient and surgical variables to aesthetic outcomes and to compare the outcomes of different kinds of breast cancer treatments (*e.g.*, reconstruction procedures). I postulated that quantitative measures of breast aesthetic properties can be designed using clinical photographs. In this dissertation, I have designed algorithms to compute objective, quantitative, reproducible measures of breast aesthetics. I have evaluated the algorithms for computing objective measures of breast aesthetic properties such as ptosis and surgical scars from clinical photographs. A preliminary observer rating scale of 11 symmetry ratings items, 14 individual breast ratings items, and a global rating on overall appearance before and after the entire rating items was proposed. Eye-tracking technology was used to understand how plastic surgeons assess breast aesthetics by recording their gaze path while they rate breast anatomy on clinical photographs. In addition to these design and evaluation tasks, I also have used the objective measures to conduct a preliminary comparison of the aesthetic outcomes of different reconstruction procedures.

Table of Contents

List of Tables

List of Figures

Chapter 1: INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

Breast cancer is the most common cancer among American women. One in eight women will be diagnosed with breast cancer during her lifetime. It is predicted that in 2007, there will be 176,295 new cases diagnosed [1]. Improved screening and treatment methods have increased the breast cancer survival rate so that the majority of women with breast cancer with localized, early stage disease can be cured when diagnosed early. Essentially all breast cancer treatment involves surgery. Breast cancer survivors face a myriad of choices to be made with the assistance of their multi-disciplinary breast care team. They must make decisions about surgical, radiation, chemotherapy, and endocrine therapy options. They must weigh concerns about survival, recurrence, co-morbidities, pain, and sexuality. They must confront these issues with incomplete and, at times, contradictory information. In addition to continuing to strive to improve detection and survival rates, more work is needed on issues that influence the quality of breast cancer survivors' lives.

The two most generally performed surgical treatments for breast cancer are breast conservation therapy and mastectomy followed by breast reconstruction. Mastectomy followed by breast reconstruction is a common form of breast cancer treatment for the patient who is in advanced stages or has contraindications to radiotherapy [2, 3]. The American Society of Plastic Surgeons (ASPS) estimates that approximately 56,000 breast reconstructive surgeries were performed in the US in 2006 [4]. Reconstruction can help breast cancer survivors regain their quality of life by restoring physical balance and helping them carry out their daily lives without being constantly reminded that they are cancer patients. Consequently, insurance providers are required to provide coverage for

all stages of breast reconstruction under the Women's Health and Cancer Rights Act of 1998 (P.L. 105-277). However, while the law requires insurance companies to provide breast reconstruction, it doesn't specify the kind of reconstruction that must be done. Healthcare providers, patients, the insurance industry, and the government all want to quantitatively assess the outcomes of breast reconstruction surgery in order to establish clinical guidelines and help breast cancer survivors make informed choices about their treatment options. A system for objective assessment is also needed in order to set charges (financial compensations) for different reconstructive procedures. Adequate measures of surgical outcome are important to protect patients' health and well-being.

Aesthetics refers to physical characteristics of the breasts, such as shape, flow, ptosis, and symmetry [5]. Accurate measurement of breast aesthetics is crucial for investigating the variables that affect the outcomes of breast reconstruction as well as other breast cancer treatments. Approaches to measuring breast aesthetics include subjective assessment by human observers [6-46], measurements on the patient (anthropometry) [8, 9, 12, 14-18, 23, 28, 47-50], measurements on photographs (photogrammetry) [9, 14, 26, 51-53], and measurements using three-dimensional images (stereophotogrammetry) [54-58]. Currently, physicians, patients, or other observers evaluate breast aesthetics in a subjective, qualitative manner. However, such assessments are typically based on vaguely defined rating scales that have low intra- and inter-observer agreement. Their qualitative nature also restricts the analyses that can be performed. Quantitative, objective measures with high reliability are needed to meaningfully relate patient and surgical variables to aesthetic outcomes and to compare the outcomes of different kinds of breast cancer treatments (*e.g.*, reconstruction procedures). The lack of a generally accepted quantitative method for assessing breast

aesthetics limits the effective assessment of the outcomes after breast reconstructive surgery.

In this dissertation, as a step toward objective assessment of aesthetic outcomes of breast cancer treatment, I develop and evaluate algorithms for computing objective, quantitative, reproducible measures of breast aesthetics factors after breast reconstruction.

1.2 OVERVIEW OF THE DISSERTATION

The dissertation is composed of a total of 7 chapters. Chapter 2 is extended from my review paper on assessment of breast aesthetics [59]. Relative to my review paper, this chapter has been updated and extended to include the detailed information about surgical breast cancer treatment methods including mastectomy followed by reconstruction. In Chapter 2, I briefly introduce the current status of breast cancer treatment, and the use of mastectomy followed by reconstruction for breast cancer treatment. Then, I review the literature on the various breast aesthetics outcome assessment methods currently used in four categorized groups: subjective ratings, physical measurements, photographic measurements, and 3-D measurements.

In Chapter 3, I propose innovative quantitative, objective measurements of breast ptosis based on ratios of distances between fiducial points manually identified in lateral and oblique views of clinical photographs. The new objective measures were compared to ratings on a subjective scale made by three experienced clinical observers. Intra- and inter-observer variability of the new objective measures and subjective ratings were investigated. In this study I found that novice observers could reliably locate the required fiducial points as compared to the identifications made by clinical observers. Moreover, both novices and clinical observers demonstrated low variability in repeated marking of

fiducial points at different time points. Over all there was good agreement between the novice and clinical observers, but some differences were seen.

In Chapter 4, I compare color measurements obtained by clinical digital photography to those from a standard colorimeter. Experimental conditions were controlled by performing measurements on artificial scars created by a makeup artist (TC). The colorimeter measurements of the artificial scars were compared to those reported in the literature for real scars in order to confirm the validity of this approach. Agreement in photographic and colorimeter measurements was statistically analyzed using the hypothesis test for equivalence, the intraclass correlation coefficient (ICC), and the Bland-Altman method. I found that photography can be used in place of a colorimeter for measuring color properties of skin in a controlled setting. I also demonstrated that artificial scars created by a makeup artist under the guidance of a surgeon exhibit color properties consistent with real scars.

In Chapter 5, I present results on using clinical photographs for the assessment of linear breast surgical scars. Digital clinical photographs were used to assess quantitative, objective measurements of breast surgical scars based on color intensity image analysis and area measurements. We demonstrate that the new objective measures utilizing clinical photographs are feasible, effective measurement methods using a hypothesis test for equivalence, the ICC coefficient, and Bland-Altman methods.

In Chapter 6, I propose an observer rating scale of 11 symmetry ratings items, 14 individual breast ratings items, and a global rating on overall appearance before and after the entire rating items were proposed. The observer rating scale was assessed for the intra- and inter-observer variability against using rating data collected by reconstructive surgeons not involved in designing the scale.

In Chapter 7, I describe eye-data collected from two plastic surgeons who viewed clinical photographs of women who underwent breast reconstruction. Mean dwell time across rating items, patients, and observers were compared. The relationships between the eye-position analyses and the surgeons' observer ratings were also analyzed. Some consistent patterns in dwell time were also seen across observers and cases. For example, both surgeons spent more time looking at the AP views than at the lateral and oblique views. This study demonstrated that eye-tracking analysis can help us understand how plastic surgeons assess breast aesthetics by using eye-tracking technology to record their gaze path while they rate breast anatomy on clinical photographs.

In Chapter 8, I summarize my findings and suggest improvement for more effective outcome assessment method using photographs or other types of imaging techniques.

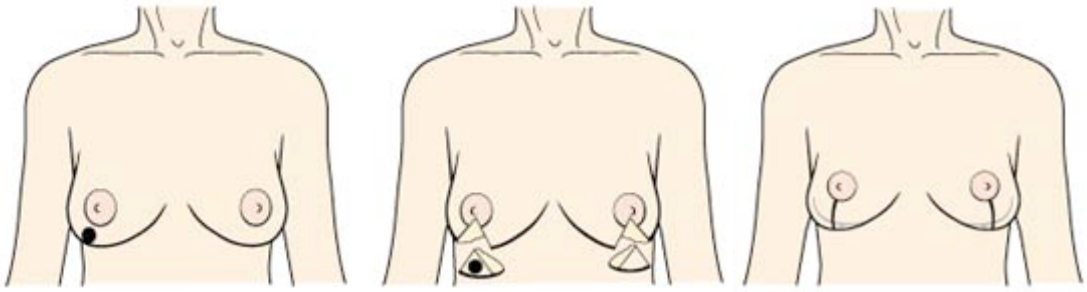
Chapter 2: BACKGROUND

***Contribution and publication**:* the study in this chapter will be published in
1. M. S. Kim, J. C. Sbalchiero, G. P. Reece, M. J. Miller, E. K. Beahm, and M. K. Markey, "Assessment of Breast Aesthetics," *Plastic & Reconstructive Surgery*, in press 2007.

2.1 BREAST RECONSTRUCTIVE PROCEDURES

2.1.1 Mastectomy

Mastectomy is a surgical cancer treatment procedure, where not only the known tumor but also some of the surrounding breast skin and tissue are removed to ensure that all cancer cells are removed. There are different forms of mastectomies that a patient may undergo depending on her cancer status: simple mastectomy, modified radical mastectomy, radical mastectomy, total mastectomy, and skin-sparing mastectomy.



(a) (b) (c)

Figure 2.1: Example illustrations of partial mastectomy. (a) The tumor in the figure's right breast is indicated by the black dot. (b) The breast segment containing the tumor is removed, and the opposite breast is reduced to match. (c) The remaining breast tissue is then reshaped into a smaller breast. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 93, Figure 21.1. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)

The difference between the mastectomy types is the amount of breast tissue removed during the procedure. For example, the nipple, areola, minimal breast skin, and all breast tissue are removed in simple mastectomy; on the other hand, the nipple, areola, and entire breast gland are removed, but all uninvolved breast skin is preserved in skin-sparing mastectomy. Variations of the mastectomy such as modified radical mastectomy involves removal of the nipple, areola, some breast skin, the breast tissue under the breast skin, and most of the lymph nodes in the axilla [3, 60-62].

2.1.2 Breast reconstruction

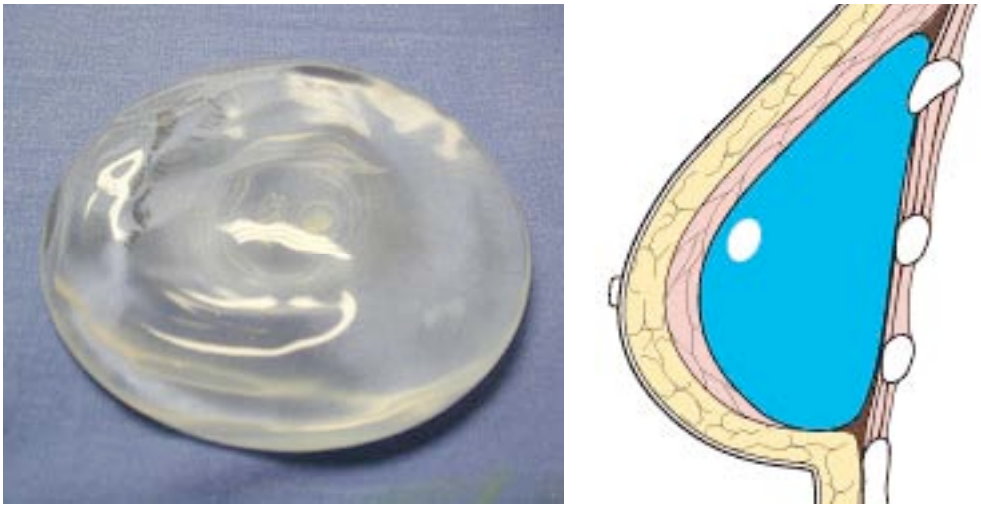
For women who must undergo mastectomy, breast reconstruction is a critical part of care after breast cancer. Breast reconstruction improves many important psychological and aesthetic benefits by physically restoring the breast or the portion of the breast removed, and helps many women to return to the life without reminding them of that they were breast cancer patients. Breast reconstruction with current technology, however, does not fully restore woman's body as it was before the mastectomy was performed. Reconstructed breast that looks very natural will not restore the partial sensation even though they may look and move naturally as natural breast does: therefore, successful breast reconstruction will make women much closer to their original physical and psychological state than women who have not had one.

A woman may undergo either delayed or immediate breast reconstruction according to her risk of cancer recurrence after mastectomy. While immediate reconstruction typically has a better aesthetic outcome, a patient has to undergo delayed reconstruction if irradiation is to follow mastectomy since radiotherapy can increase complications after surgery [63, 64].

Decisions about reconstructive surgery will depend on many factors such as: patient's overall health, stage of breast cancer, size of breast, amount of breast tissue, and personal preference. Two basic types of breast reconstructions are performed: implant-based, where artificial material is replaced with missing breast and tissue-based, where living material, mainly skin and fat from patient's own body are used [5].

In implant-based reconstruction, symmetry, one of the most important aesthetic factors, is easier to achieve for bilateral reconstruction rather than for unilateral reconstruction. However, an additional procedure may be necessary for unaffected breast, called mastopexy (breast lift) to obtain symmetry between breasts and the created symmetry tends to be temporary since the natural breast becomes more ptotic (breast droopiness) with time. Moreover, a breast reconstructed with an implant usually does not have natural flow (the upper mound becomes convex unlike natural breast which is concave) [65].

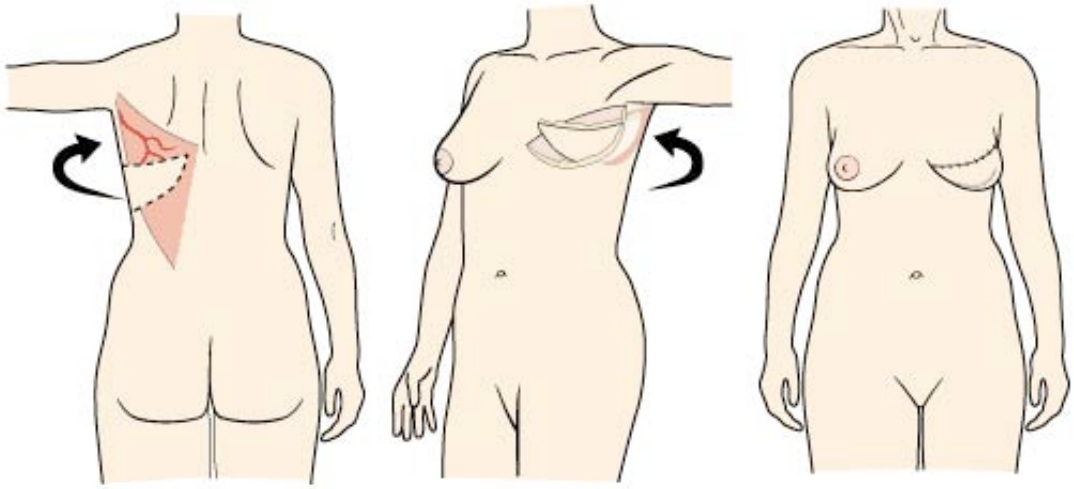
The main advantage of tissue-based reconstruction is that affected breast will behave like a real breast (weight, natural ptosis, and feel). Moreover, the symmetry normally lasts longer compared to that of implant-based reconstruction. There are three most popular tissue-based reconstructions to be discussed: LD (latissimus dorsi), TRAM (transverse rectus abdominis myocutaneous), and DIEP (deep inferior epigastric perforator) [65, 66].



(a) (b)

Figure 2.2: Illustration of implant-based reconstruction. Cross-section of a breast reconstructed using an implant shows the position of the implant beneath the muscle and skin of the chest wall. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 16-17, Figure 4.1, 4.2. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)

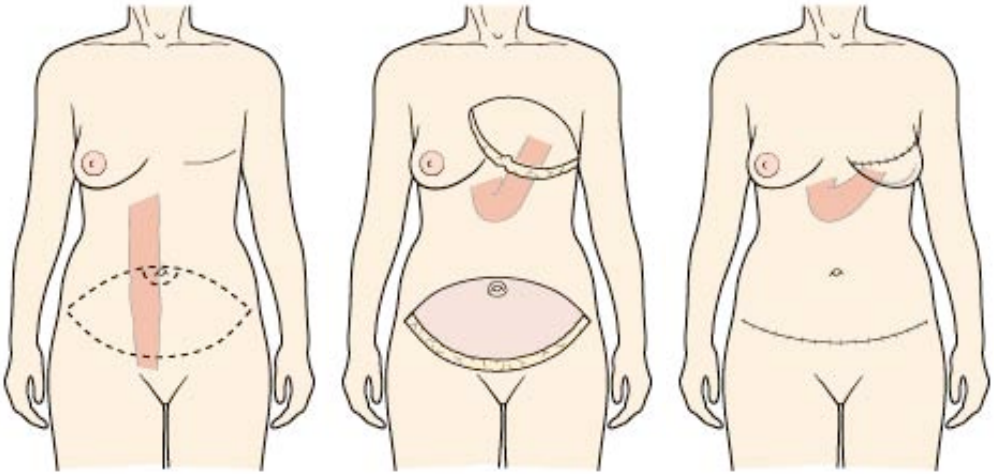
In LD reconstruction, the flap composed of an elliptical patch of skin that lies over the latissimus dorsi muscle along with the muscle are replaced with breast skin when the skin was removed during and the breast volume is replaced with a breast implant in most cases. Breast implants are less visible in LD procedure than tissue expansion, which results in a more natural appears than implant-based reconstruction with tissue expansion [65].



(a) (b) (c)

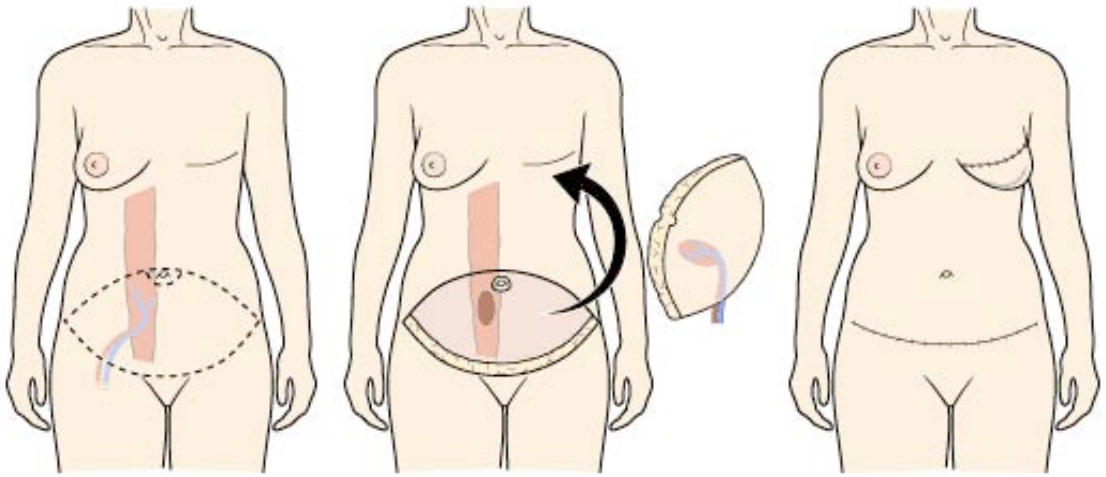
Figure 2.3: Illustration of reconstruction with latissmus dorsi (LD) muscle flap. (a) The plan for an LD flap shows the elliptical patch of skin overlying the latissimus dorsi muscle. (b) The flap is separated from the body except at the point where blood vessels enter the muscle. The flap is passed under the axillary skin into the breast area. (c) The LD flap is usually used to cover a breast implant. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 28, Figure 7.1. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)

The transverse rectus abdominis myocutaneous (TRAM) is the most popular tissue-based breast reconstruction, and two most common types of techniques are performed: TRAM pedicled flap [67-71] and the TRAM free flap [72]. In TRAM, the flap either with blood vessels either remained intact or completely severed from the body is transferred to the chest via a subcutaneous tunnel to replace the breast skin and volume. In TRAM free flap, breast can be shaped effectively with flap over pedicled flap. In both types of TRAM, the umbilicus remains intact but offset [65, 68, 73-75].



(a) (b) (c)

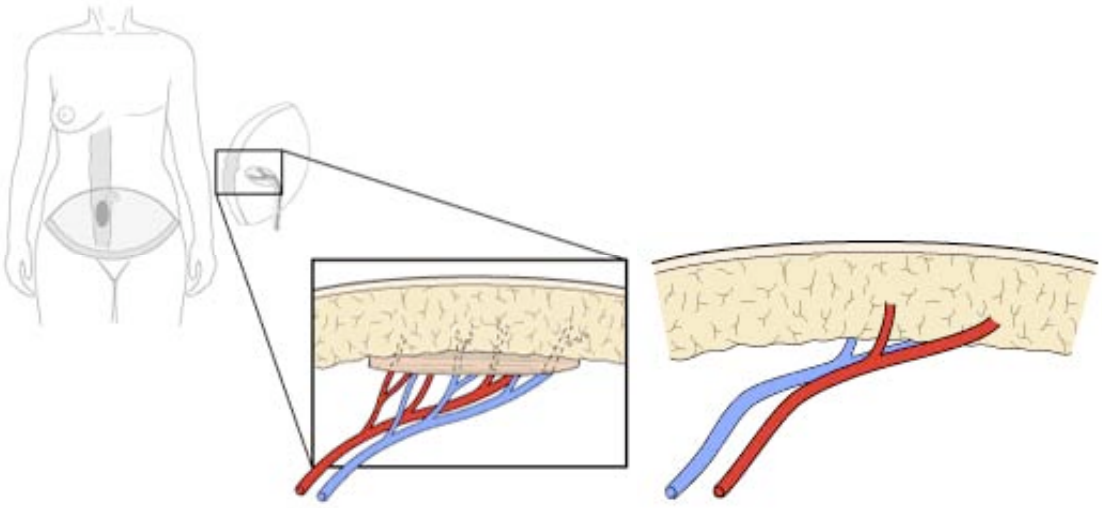
Figure 2.4: Illustration of breast reconstruction with a TRAM pedicled flap. (a) The TRAM pedicled flap contains skin, fat, and underlying muscle from the abdomen. This illustration shows the plan for the surgery (dotted line) with the muscle and blood vessels beneath the skin and fat. (b) The flap is passed through a "tunnel" beneath the upper abdominal skin and into the mastectomy area.(c) Blood is supplied from blood vessels in the rectus abdominis muscle. The flap is shaped into the new breast. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 34, Figure 8.2. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)



(a) (b) (c)

Figure 2.5: Illustration of breast reconstruction with a TRAM free flap. (a) A plan for a TRAM free flap shows the muscle and blood vessels. (b) The flap and vessels are separated from the body. The blood vessels are preserved for connection to blood vessels in the chest or axilla areas. (c) The blood vessels are connected to restore circulation to the flap, which is then shaped to form the new breast. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 36, Figure 8.3. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)

The DIEP free flap is a variation of the TRAM flap and is considered by plastic surgeons to be a type of TRAM free flap. The flap therefore consists of skin, fat, and the blood vessels but, unlike a TRAM flap, contains no muscle [76].



(a) (b)

Figure 2.6: Illustration of breast reconstruction with a TRAM free flap. (a) The standard TRAM free flap contains a portion of the rectus abdominis muscle along with skin, fat, and blood vessels. (b) In contrast, the DIEP free flap contains no muscle - only skin, fat and blood vessels. (Reprinted, by permission, from Reece GP. Breast reconstruction guide for patients: The well-informed patient guide to breast reconstruction, p 40, Figure 9.1. © 2002 by The University of Texas M. D. Anderson Cancer Center, Houston)

2.2 ASSESSMENT OF BREAST AESTHETICS

The goal of plastic surgery of the breast is to recreate a natural appearance that is satisfying to the patient. Aesthetic breast surgery is intended to alter the appearance of a normal breast to more closely approach some preferred aesthetic ideal as understood by the patient. Reconstructive breast surgery has the identical purpose but starts with a breast considered abnormal in appearance due to a deformity or complete absence as following total mastectomy. Despite the fundamental nature of this outcome, breast aesthetics is poorly understood and difficult to measure. In this context, aesthetics refers to visible physical characteristics of the breast that determine the subjective evaluation of overall appearance [5]. Relevant factors include breast size, shape, proportion, ptosis, symmetry,

skin quality, and nipple location [77-81]. Although these are generally understood as important, there remains no objective, quantifiable determinants of breast aesthetics that are universally acknowledged [82]**.** This absence of standards has several undesirable consequences. It makes it difficult to compare outcomes of different surgical techniques. It also fosters discord when a discrepancy exists between the aesthetic impression of the plastic surgeon and the patient. Significantly different perceptions of body image regarding attractiveness, natural, and ideal breast shape remain among patient and plastic surgeon groups [83, 84]. The relationship between physical appearance and psychological body image is a central aspect of a patient's quality of life, and study of this relationship can only be achieved through appropriate aesthetic outcome assessment [85-90]. It is therefore essential to characterize the elements of aesthetic breast appearance in order to make meaningful comparisons between competing surgical techniques, to adequately inform patients of realistic outcomes, and to make systematic advances in treatment. Objective measurement will reduce the gaps between the plastic surgeon's surgical success and patient's desire. In addition, objective measures are crucial for documenting to third party payers and health policy analysts the clinical value of costly techniques such as autologous breast reconstruction [81, 91, 92].

Current approaches to breast aesthetic measurements include: subjective assessments by human observers, measurements on the patient's body (anthropometry), measurements on photographs (photogrammetry), and measurements using three-dimensional images of the breasts.

2.2.1 Observer rating scales

Breast aesthetics is commonly assessed by visual assessment, which typically employs a crude gradation scale and is inherently subjective and qualitative (Table 2.1).

Substantial variability has been reported in the use of common scales of four gradations of aesthetic change ($\kappa = 0.21 - 0.31$) [93, 94]. Likewise, the reported reliability of visual analogue scales is unacceptable ($\kappa = 0.13 - 0.15$)[94].

It has been argued that global scales suffer from vague terminology and that more reliable aesthetic assessments will require separate measurements of individual factors. In particular, subscales with more explicit criteria for each aesthetic component (e.g., volume, contour) have been recommended [94]. However, substantial observer variability was still reported when such subscales were employed ($\kappa = 0.19 - 0.63$) [94]. Similarly, subscales of four gradations have been proposed for rating the difference between treated and untreated breasts in terms of size, shape, skin color and firmness, and the visibility of surgical scar, but low to moderate reliability were again reported ($\kappa = 0.24 - 0.40$) [95].

It is also noteworthy that there is a lack of concordance between different observer groups [96-98]. Sneeuw *et al*. reported that patients demonstrated significantly greater variation ($\kappa < 0.10$) in aesthetic scoring compared to professionals ($\kappa = 0.64$) using global 4-point subscales, suggesting that accurate measures may require that patient assessments be combined with those of clinicians [95].

Table 2.1: Observer Rating Scales Assessment. BCT and BRC denote breast conservation therapy and breast reconstruction, respectively.

Reference	Scale	Test Population	Findings
Pezner [23]	Two 4-point ordinal scales	14 Photographs of BCT patients	Experienced observers had higher agreements than the inexperienced observers. The scales showed low reliability.
Lowery [22]	4-point, subscales (volume, contour, inframammary fold, scars), & visual analogue scales	50 photographs of BRC patients	Suggested explicit criteria and to separate various components of the aesthetic result to improve the reliability of the assessment.
Sneeuw [28]	4-point scale with subscales (scar, size, shape, color, firmness)	76 Photograph of BCT patients	Reported higher intra-rater agreement but lower inter-rater agreements between ratings by patients and clinical observers.
Cohen [99]	Questionnaire on breast aesthetics with 5-point scale	36 photographs of BRC patients	Reported better internal consistency and more reliability of ratings by patients evaluating their own results; Ratings by surgeons were not as internally consistent and reproducible. Low inter-observer agreement was observed.

In contrast, a study by Cohen *et al*., which employed a global 5-point scale, reported greater internal consistency among patients ($Chronbach\ \alpha = 0.92$) compared to medical professionals ($Chronbach\ \alpha = 0.74 - 0.89$) assessment of the breast aesthetic appearance using photographs of 36 autologous breast reconstruction patients. There was

poor inter-observer agreement among surgeons $(\kappa = 0.0 - 0.39)$ and a weak correlation between surgeons and patients $(Spearman\ \rho = 0.36 - 0.53)$ [100].

These studies highlight two major disadvantages of observer rating scales. The first is variable internal consistency and reproducibility as demonstrated by low intra- and inter- observer agreement (Table 2.1). In an attempt to overcome this problem, data averaged from a panel rather than individual evaluators are often employed, but this approach is time and labor intensive. Also, calculating an average can reduce variability but not improve accuracy. The second problem is the lack of a standardized, explicit scale for analyzing aesthetic outcome. A crude ordinal scale with four or five categories is imprecise for identifying individual aesthetic elements. Quantitative, objective measures with high reliability are needed in order to meaningfully relate patient and surgical variables to aesthetic outcomes.

2.2.2 Physical measurements

Anthropometry is based on linear measurements between surface landmarks such as the sternal notch, nipple, inframammary fold, and lowest point of the breast mound (Table 2.2). Penn's approach of defining nipple-to-sternal notch and midclavicular point distances of "aesthetically perfect" breasts of 20 women gained broad attention and has been adopted by many as normative [101]. In a similar attempt to establish normal values, the distances between fiducial points were computed in 66 women in whom one-third had either breast hypertrophy, ptosis, or both [102].

Table 2.2: Physical Measurements.

Reference	Test Population	Findings
Penn [103]	20 aesthetically normal breasts	Linear measurements between fiducial points were compared and tabulated.
Smith [47]	66 patients with breast hypertrophy, ptosis, or both	Symmetry data for right and left breasts are described.
Smith [18]	40 patients with breast asymmetry	Weak correlation was reported for nipple position relative to sternal notch, midline, and axilla in relation to observer rating score.
Westreich [49]	50 women with aesthetically perfect breasts	Of 22 linear measurement performed, nine measurements were shown to have significant correlation with breast volume.
Tsouskas [8]	151 women (100 normal breasts and 51 BCT patients)	Reported a correlation between the difference of the Breast Compliance Evaluation (BCE) measure of the two breasts and observer rating assessment.
Stark [17]	72 women with asymmetrical breasts	Breast symmetry was assessed by three linear measurements between fiducial points and the overall result was compared with patient's own evaluation.
Hauben [104]	37 women with normal breasts	Reported correlation of patient characteristics (age, height, weight, Body Mass Index (BMI)) and breast proportions and significant positive correlations between age and areola-breast proportion.

In another study, the same group measured the symmetry of 40 patients by computing differences between the right and left breast. They studied the relationship between observer ratings and linear measurements with scatter diagrams of linear measurements and aesthetic score [105]. "Aesthetically perfect breasts" was defined as a breast shape for which no aesthetic procedure would be indicated. The problem with this

definition is that different surgeons have different notions of when an aesthetic procedure is indicated. For example, in general, European Plastic Surgeons feel that what American Plastic Surgeons consider to be the "ideal breast" is really too large and they would recommend a breast reduction. Twenty two linear measurements were designed and compared to the results with those of Penn [101] and Smith [102]. Nine of the measures were shown to have statistically significant correlation ($Pearson\ r = 0.459 - 0.592$) with breast volume [49].

Symmetry has typically been determined by calculating the difference in measured distances on the breast mound and nipple/areola complex. A good example of measuring the symmetry is demonstrated in the Breast Compliance Evaluation (BCE) method, where the distance from the point of mid-clavicle to the inframammary fold through the center of the nipple was measured in both supine and erect positions using a tape measure. Compliance was calculated from controls and cancer patients by determining the difference in the measures between the two positions and the aesthetic outcome of treated breasts was evaluated by comparing the difference between the compliances of the two breasts. Correlation between BCE and satisfactory aesthetic outcomes was reported, but the reliability of the measurement was not clear [106]. Combination of physical measurement with observer ratings was attempted to measure the symmetry by making three measurements on 72 patients with asymmetrical breasts, but the correlation between the measures and reproducibility was not analyzed [107]. Areola diameter in addition to the distances between sternal notch-to-nipple, and nipple-to-inframammary fold were measured, and compared with the observer ratings for patients with developmental asymmetry [108].

Another study computed the distances between the nipple and the borders of the breast to calculate the nipple-areola-breast proportion and showed the correlation of

patient characteristics (age, height, weight, Body Mass Index) and breast proportions. Significant positive linear correlations between age and areola-breast proportion ($Pearson\ r = 0.47$) were reported and areola-nipple proportion was shown to be significantly larger in higher ptosis grades [109].

Anthropometry can be a useful tool for quantifying aesthetic outcomes. However, this method has several pragmatic limitations. Fundamental parameters such as breast projection are difficult to evaluate because of the curvature of the underlying chest wall and mobility of subcutaneous tissue. In addition, anthropometry can be complicated and time intensive, making it impractical for routine use. Studies of linear measurements on a patient must be done prospectively and they require an extra intervention, which tends to keep the number of subjects low. It is not feasible to make a large number of measurements on each subject. If a particular measurement doesn't prove valuable, one can't retrospectively try a different one. To prove the validity of a specific measurement, studies across multiple institutions with multiple observers are needed, which are costly. The relationships to observer rating scales is unclear [105, 107, 108, 110, 111]. For these reasons, direct measurements have had limited utility in routine clinical practice and are not generally performed.

2.2.3 Photographic measurements

The value of photographs for observer rating assessment of breast aesthetics was confirmed by comparable results for assessments based on photographs compared to physical examination [26]. Moreover, prints produced from digital images, digital images displayed on computer monitor, or conventional photographs have been shown to be acceptable to observers for observer rating assessment of breast aesthetics [112].

Table 2.3: Photographic Measurements.

Reference	Method or Scale	Subject	Test Population	Findings
Pezner [51]	Breast Retraction Assessment (BRA)	Patient & clear acrylic grid and photographs	29 Normal women compared to 27 BCT patients	BRA was significantly greater for breast cancer patients than for the control groups. BRA calculations correlated with the size of resection.
Van Limbergen [52, 53]	BRA and panel scoring	Anterior Posterior (AP) Photographs	142 BCT patients	Significant correlations between BRA and the observer rating scores were found. Increased BRA associated with poorer aesthetic outcome.
Vrieling [21]	BRA and pBRA	AP Photographs	647 patients	Found a significant correlation between *p*BRA and observer rating assessment at a three-year follow-up except for those with inferomedially located tumors.
Sacchini [9]	4 measures of distances between fiducial points	AP Digital images	148 BCT patients	A 3-member panel, comprised of either health care professionals or patients, assessed the images. Significant difference in the aesthetic outcome among groups was reported.
Kim [113]	Ratios of distances between fiducial points	digitized/digital images of oblique and lateral (pre-operative)	52 BRC patients	The variability in the objective measurements due to intra- and inter-observer variability in marking fiducial points was shown to be equivalent to less than one point on the observer rating ptosis scale.

Rather than simply obtaining observer ratings based on photographs, several investigators [26, 40, 114-118] describe measurements calculated on digital/digitized photographs (Table 2.3). This approach has yielded encouraging results, but there are

limitations to the reported studies. The majority of these have only been reported in breast cancer patients undergoing breast conservation therapy (BCT) and those have limited utility for full aesthetic assessment. They rely on many of the same fiducial points described above for anthropometry, but some anatomical landmarks may not be visible in photographs (e.g., the inframammary fold). They also rely heavily on the anatomy of the nipple/areola complex, which may not reflect independent features of the breast mound. Finally, they typically involve only a single anterior-posterior (AP) photograph for each patient.

The Breast Retraction Assessment (BRA) has been successfully applied in the assessment of asymmetry [117]. The BRA measure is the Euclidean distance between right and left nipples. Early studies employed the aid of a clear acrylic sheet marked as a grid with the patient behind the screen, which is functionally equivalent to calculating distances on a digitized/digital photograph. BRA was compared between normal volunteer and 27 BCT patients, and the measure was significantly greater for the breast cancer patients than for the control group. BRA calculations correlated with the size of resection [117].

Van Limbergen *et al.* [114, 115] calculated four measurements on AP photographs: the vertical distance from the level of the sternal notch to the nipple (A), the vertical distance from the level of sternal notch to the inferior pole of the breast (I), the horizontal distance from the midline to the nipple (M), and the horizontal distance from the midline to the lateral breast contour (L). The authors used the differences in each measure between the left and right breasts as measures of symmetry (*e.g.*, ΔI).

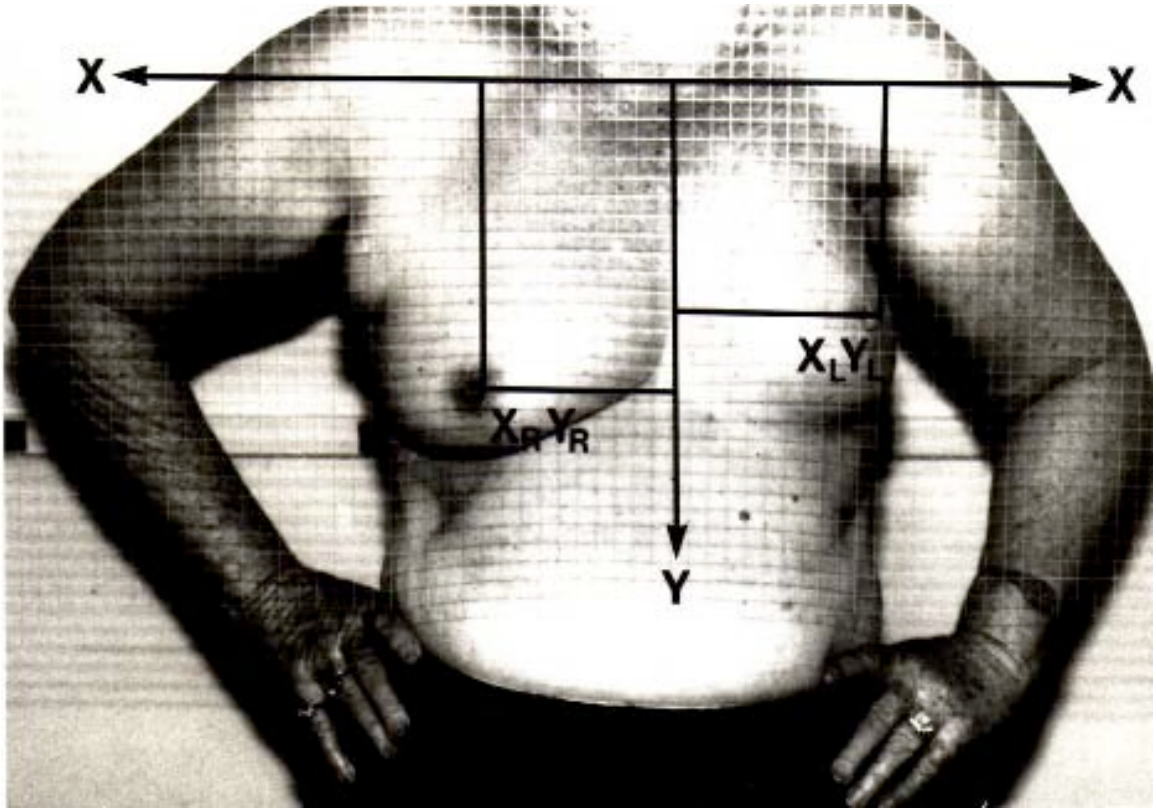


Figure 2.7: Breast Retraction Assessment (BRA). AP view of the patient standing behind acrylic grid plate. X- and Y-axes are marked on the photographs and distances between nipples are calculated using the grids for calculating BRA. *"Reprinted from Publication* Pezner, R.D., et al., *Breast retraction assessment: an objective evaluation of aesthetic results of patients treated conservatively for breast cancer*. International Journal of Radiation Oncology, Biology, Physics, 1985. **11**(3): p. 575-8. *with permission from Elsevier"*

Observer rating assessment by a panel comprised of a surgeon, a radiotherapist, and a non-physician was also made from the photographs using a five-point scale and the mean panel rating was correlated to the BRA measure. In 142 BCT patients, significant correlations between BRA and the Observer rating scores ($p = 0.0001$: *X square test, Kendall Tau B and Kendall Tau C*) were found with increased BRA associated with poorer aesthetic outcome. Since the magnification of clinical

photographs is not standardized, they introduced a relative version of the BRA measure (*p*BRA) equal to BRA divided by the distance from the sternal notch to the nipple of the untreated breast. Using this measure they found smaller intra- and inter-observer variability based on *p*BRA than on observer rating assessments. They also found a significant correlation between *p*BRA and Observer Rating assessment at a three-year follow-up except for those with inferomedially located tumors ($\rho = 0.24 - 0.53$) [115, 116] (Figure 2.1).

Some studies have employed a computer program to obtain more consistent measurements [118]. Four measurements were calculated on AP images of 148 women who underwent BCT. Distances were computed for from each nipple to the contralateral nipple, the inferior pole of the breast, the midline, and the sternal notch. A 3-member panel comprised of either health care professionals or patients then assessed the images. Significant differences were reported in the aesthetic outcome between each method of assessment.

Quantitative and objective measurements of breast ptosis have been designed based on ratios of distances between fiducial points manually identified in digitized/digital images of oblique and lateral clinical photographs (pre-operative). These measures were compared to ratings made using a 4-point scale by Regnault [119]. The new objective measures of ptosis showed encouraging levels of concordance with ratings made by experienced surgeons using the observer rating scale. The objective measures were found to be robust to intra- and inter-observer variability in marking fiducial points, including identifications by "novice" observers. The variability in the objective measurements due to intra- and inter-observer variability in marking fiducial points was shown to be equivalent to less than one point on the observer rating ptosis scale [120].

Photogrammetry has advantages over anthropometry. A photograph is more efficient and less intrusive for the patient. They create a permanent record from which it is possible to make a variety of measurements. Digital/digitized photographs can be analyzed using an automated process on a computer, potentially yielding more consistent objective results. The disadvantages are that some anatomic landmarks may not be visible and the measurements cannot be obtained following the contours of the patient's body. Some studies have reported substantial intra- and inter-observer standard deviation for linear measurements on photographs, related primarily to the lack of consistency in the manual identification of anatomic landmarks [116]. Consistent guidelines for standard photography are critical to obtaining reproducible assessment of aesthetic outcomes from one institution to another [121-125].

There are commercial software systems for working with digital/digitized photographs in plastic surgery, such as Nautilus Plastic Designer™ (NauSoft LLC, St. Louis, MO), Mirror™ software (Canfield Clinical Systems, Fairfield, NJ), and iMARS™ software (iMARS Medical Office Management Systems). However, it should be noted that most products are focused on data management rather than analysis. Some such systems will allow the surgeon to simulate or morph the postoperative result in an artistic manner. However, these simulated images are not really representative of the actual surgical outcome.

2.2.4 Three-dimensional imaging

Three-dimensional (3D) imaging (Figure 2.2, Table 2.4) is a recent innovation applied to assessing breast appearance. Several technologies such stereophotogrammetry, laser scanning, three-dimensional digital photography, and light digitizers are used to create 3D images that have advantages in the analysis of human three-dimensional

structures [126]. This utilization is well documented in craniofacial and facial plastic surgery for aiding in surgical planning, evaluation of facial soft tissue volume changes and symmetry, and in objectively assessing the results [126-131] .

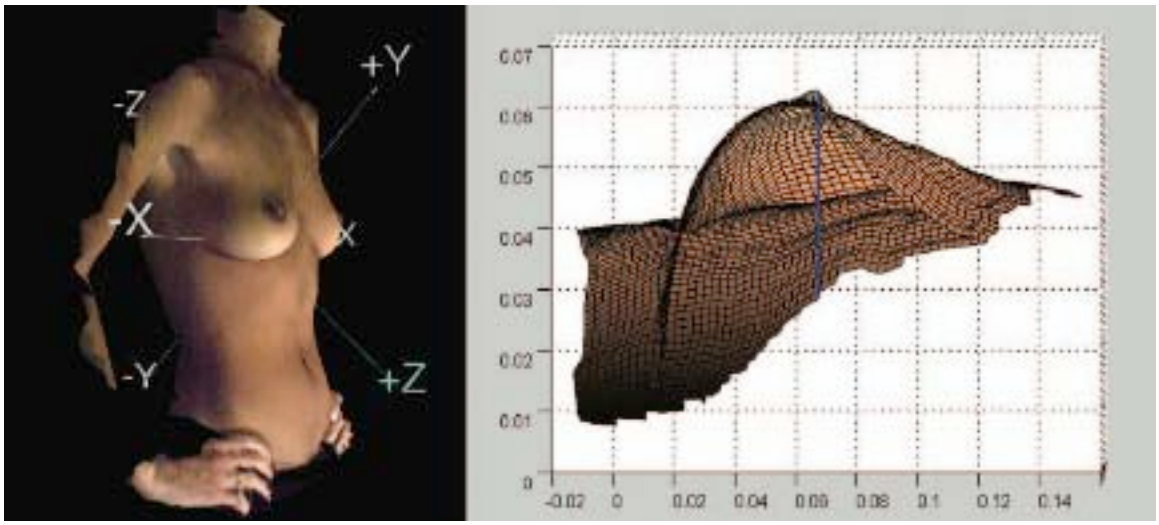


Figure 2.8: Representative images illustration and measurement capabilities. Left image displayed laterally rotated image and co-ordinate axes. Right view displayed using the mesh mode and surface of the breast. Breast projection and volume enclosed between the base and the surface are quantitatively estimated. Reproduced images by 3dMDtorso Imaging System (3-Q, Technologies Ltd., Atlanta, GA).

Three dimensional digital photography systems are capable of non-invasively and quickly capturing high-definition 3D images and constructing topographic surface maps of the breast that permit accurate evaluation and objective determination of differences in volume, surface area, shape, size, contour, and symmetry [57, 127, 132-134]. A single 3D image yields more information regarding breast appearance than multiple conventional photographs including data regarding some elements of the breast appearance, such as volume, that are not available from two-dimensional images [82]. Assessment of the

degree of soft-tissue edema following surgery is another important application of 3D imaging [57, 135].

Breast volume is an extremely useful parameter for the correction of breast asymmetry in breast reconstruction. For example, for a patient seeking delayed unilateral reconstruction, the calculated volume of the contra-lateral breast could facilitate the selection of implant/tissue volume that would provide improved symmetry. Many methods have been described to determine breast volume: water displacement, the Grossman-Rounder device, thermoplastic casting, biostereometric analysis, and radiographic techniques [100, 136-144]. Most of them are inaccurate, time-intensive, or expensive. Digital 3D imaging overcomes these limitations, permitting automated calculation of breast volume. Both semi-automated [145] and automated image analysis [146] methods have been developed for measurement of breast volume from 3D data. Once an image is acquired, it is possible to rotate the subject on the computer screen to see not only the front and side views, but also at any angle in between [147]. Point-to-point measurements occur in a tangential arc to the surface topography rather than in conventional linear measurements. Thus, volume, surface distance and area are ascertained by measurement of soft tissue overlying the bitmap wire frame of the digital image.

Galdino *et al*. analyzed 3D images of over 50 patients who underwent breast reconstruction out of over 100 patients who underwent 5 types of breast surgeries [57]. They used the images to preoperatively estimate the expander and implant volumes and decide the type of surgery between TRAM versus DIEP, and to postoperatively assess breast asymmetry for planning subsequent revisions.

Table 2.4: Three-dimensional Imaging.

Reference	Test Population	Findings
Galdino [57]	Over 100 patients of 5 types of breast surgery	Extremely large breasts need additional manual intervention to capture the image. Volume of breasts with significant ptosis can be overestimated (contrast with the observation by Loosen [148]).
Kovacs [149]	1 BRC patient	Breast volume was easily calculated in the virtual models as done with thermoplastic casts. Some technical limitations were observed
Loosen [148]	14 BRC patients	Difficulty was experienced in the analysis of very ptotic breasts, with a tendency to under calculate breast volume (contrast with the observation by Galdino [57]). Inherent difficulties in consistently identifying breast boundaries were observed

Kovacs *et al*. demonstrated that three-dimensional imaging contralateral breast volume can be used to estimate volume requirements for breast reconstruction [82]. The accuracy and reproducibility of 3D analysis was tested by comparing breast volumes calculated from 3D images to actual volumes measured from mastectomy specimens by water displacement. Difficulty was experienced in the analysis of very ptotic breasts, with a tendency to under calculate breast volume. Variability in volume analysis was attributed to difficulties in consistently identifying breast boundaries. A more useful and accurate indication for 3D imaging of the breast would be to evaluate longitudinal changes in breast size, volume, shape, and surface area by superimposing images over time and calculating differences with accuracy [135].

Commercially available 3D imaging systems for plastic surgeons include features such as morphing operations to alter sizes (such as breast sizing for mammaplasty), and a broad range of art tools for simulation of aesthetic alterations on images. Some of the popular software systems include 3dMD Breast Analysis™ (3-Q, Inc., Atlanta, GA) and 3D Surgeons (Genex Technologies Inc., Kensington, MD). However, as is the case for systems for analyzing 2D images, some current products allow unrealistic simulations of surgical outcome that are based on very little or no quantitative information on specific breast aesthetic parameters (*e.g.*, symmetry).

Three-dimensional imaging has tremendous potential for analysis of breast appearance. The technique does have limitations, particularly for women with large, ptotic breasts and patient positional changes do influence measurements. Nevertheless, 3D systems may offer the most accurate means to quantify numerous elements of breast appearance and evaluate changes over time [57, 150]. Further development of this technology might yield a variety of useful clinical tools to aid surgical planning, patient decision making, and outcomes analysis [82, 147, 150].

2.3 SUMMARY

The contemporary goals of breast cancer treatment are not limited to cure but include maximizing quality of life. All breast cancer treatment has the potential to adversely affect breast appearance and cause morbidity due to deformity. Developing objective, quantifiable methods to assess breast appearance is important to understand the impact of deformity on patient quality of life, guide selection of current treatments, and make rational treatment advances.

Currently, there is no reliable means to meaningfully assess breast appearance. Current methods are plagued with poor reproducibility and are excessively influenced by

observer rating interpretations. Those that yield more objective data are based on physical measurements, but they are time and labor intensive and generally impractical for routine clinical use. A more complete understanding of how these objective measures are related to qualitative aesthetic judgments is also needed.

Aesthetics will always be an elusive outcome due to the variability of culture, ethnicity, and personal preferences. Nevertheless, the physical elements that define breast appearance do permit objective evaluation and may serve as tools to gain better control over aesthetic outcomes. The most important issue in breast reconstruction seems to be achieving symmetry with the contralateral breast even when it is not an aesthetically pleasing breast. There is a division between anatomic parameters that determine breast appearance and the subsequent observer rating judgments related to aesthetics, but these are interrelated concepts. Developing tools to more objectively assess the anatomic components of breast appearance will contribute to better evaluation of aesthetics. More research is needed to develop an evaluation method with high reliability and reproducibility that is also practical. A detailed observer rating qualitative scale with standardized terminology would improve communication between reconstructive surgeons and radiation oncologists. Quantitative calculations based on standardized digital/digitized two-dimensional photography could be a reliable and reproducible approach that takes advantage of existing, easy access technology. Three-dimensional imaging has even greater potential, though more substantial development is needed.

In the absence of an absolute “gold standard”, a comparative analysis of quantitative aesthetic outcome assessment by medical professionals, assessment by patients, and degree of patient satisfaction would be useful for identifying factors that influence a woman’s quality of life after breast cancer treatment and rehabilitation and as a check of the appropriateness of new assessment measures. Highly reliable outcome

scales would allow more meaningful comparison of results and effectiveness among surgical techniques. They would be applicable not only to breast cancer treatment but to all forms of plastic surgery of the breast.

Chapter 3: MEASURING PTOSIS FROM CLINICAL PHOTOGRAPHS

***Contribution and publication**:* the study in this chapter was published in the journal of
1. M. S. Kim, G. P. Reece, M. J. Miller, E. Beahm, E. N. Atkinson, and M. K. Markey, "Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," *Computers in Biology and Medicine*, vol. 37, pp. 49-59, 2007. and was presented at the conference
2. M. S. Kim, G. P. Reece, E. N. Atkinson, and M. K. Markey, "Objective assessment of the aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," presented at BECON/BISTIC 2004 Symposium on Biomedical Informatics for Clinical Decision Support: a Vision for the 21st Century, Washington D.C., 2004.

3.1 INTRODUCTION

An important aesthetic property that has not been addressed by prior photogrammetry studies is ptosis. Ptosis of the breast is classified according to the relationship between the nipple and the inframammary fold, the crease beneath the breast. Unacceptable ptosis occurs when nipple and lower pole of the breast descend lower than the level of inframammary fold [15]. Regnault [16] defined four grades of ptosis. A patient has no ptosis when the nipple and most of the breast gland is located above the level of submammary fold (Grade 0). In first degree or minor ptosis, the nipple is at the level of submammary fold and above the lower contour of the breast (Grade 1). Second degree or moderate ptosis is when the nipple lies below the fold but above the lower contour (Grade 2). Third degree or major ptosis is when the nipple lies at the lower breast contour and most of the breast is below the fold (Grade 3). The ptosis scale with example photographs is described in Table 3.1 and Figure 3.1.

In this study, I designed quantitative, objective measurements of breast ptosis based on ratios of distances between fiducial points manually identified in lateral and oblique views of clinical photographs. The new objective measures were compared to ratings on an observer rating scale made by three experienced clinical observers. Intra-

and inter-observer variability of the new objective measures and Observer Rating ratings were investigated.

Table 3.1: Observer rating ptosis scale based on Regnault [16] and Bostwick [15]. Sample images are shown in Figure 3.1.

Ptosis Grade	Definition
0 = None	Nipple and most of gland are above IMF
1 = Minor	Nipple at IMF
2 = Moderate	Nipple is below IMF but above lower contour of breast
3 = Major	Nipple at lower breast contour & breast below IMF

Grade 0 Grade 1 Grade 2 Grade 3

Figure 3.1: Sample images (right lateral view) illustrating the ptosis grades described by Regnault [16] and Bostwick [15]. *n*: nipple point, *i*: lateral terminus of inframammary fold.

3.2 MATERIALS AND METHODS

3.2.1 Datasets

The patient population for this study was women aged 21 years or older who underwent breast reconstruction surgery from January 1, 1990 to June 1, 2003. Anterior-posterior (AP), right and left lateral, and right and left oblique photographs were digitally taken with a Nikon 990 Coolpix or Cannon T90 35mm SLR with 50mm lens and digitized with a Nikon LS 2000 or Nikon Super Coolscan 4000ED (1.06) slide digitizer. An experienced plastic surgeon (GPR) selected pre-operative images for 52 patients. The dataset demonstrates a wide range of breast aesthetic characteristics including size, symmetry, ptosis, and projection.

3.2.2 Ptosis: observer rating

The ptosis of each breast of all 52 patients was rated by three clinical observers (GPR, MJM, EKB) independently using a 4-point scale (Table 3.1, Figure 3.1) described by Bostwick [15]. All five views of a patient were displayed simultaneously. The observer rating assessment was repeated at three time points at least 2 weeks apart. Thus, it is unlikely that an observer recalled his/her previous rating. Kappa statistics of observer rating scale measures between time points (1^{st} vs. 2^{nd}, 2^{nd} vs. 3^{rd}, and 1^{st} vs. 3^{rd}) were analyzed to investigate intra-observer variability. Kappa statistics of the observer rating scale measures between clinical observers (GPR vs. MJM, GPR vs. EKB, and MJM vs. EKB) were analyzed to investigate inter-observer variability.

3.2.3 Fiducial points

Out of the 52 patients whose names were unknown to any observer, a clinical observer (GPR) selected ten patients that illustrate a wide range of aesthetic characteristics (size, ptosis, and projection) and ethnicity. Eight observers (three clinical (GPR, MJM, EKB) and five novices) marked fiducial points in images of the ten patients on three occasions with 5 minutes breaks between assessments. The novice observer group was trained to mark fiducial points by a clinical expert (GPR) before the study was performed using a separate set of photographs. There were no additional guides or hints provided for marking fiducial points during the study. The four fiduical points marked were the sternal notch, lateral extent of the inframammary fold, the lowest visible point, and the nipple centroid. The sternal notch (*s*) or jugular notch point is the shallow indentation on the superior surface of the manubrium. It is located between two clavicular articulations [15, 17]. The inframammary fold is a curvilinear structure which is generally hidden behind breast tissue since most women have some degree of ptosis [15]. The lateral terminus of the inframammary fold is the endpoint of the inframammary fold (*i*) where it intersects at the chest wall and it is typically visible in the lateral and oblique photographs. Therefore, the lateral terminus of the inframammary fold was used as the reference point. The lowest visible point (*v*) of the breast is located at the most inferior point of the breast. The centroid of nipple (*n*) is considered rather than the nipple-areola complex since many women have an irregular shaped areolae. The sternal notch (*s*), lateral terminus (*i*) and the nipple (*n*) are marked in the oblique views of the right and left breasts separately. The nipple (*n*), the lowest visible point (*v*), and the lateral terminus (*i*) are marked in the lateral views of the right and left breasts separately. We designed a program using MATLAB® (The MathWorks, Natick, MA) to automatically load the images and prompt the observer to manually identify locations of the fiducial points. The

x- and y- coordinates and the time required per each fiducial point were recorded. We studied x- and y- localizations for the fiducial points for both clinical and non-clinical observers. Minimum and maximum values of variability of the fiducial points for both groups were calculated.

3.2.4 Ptosis: objective measures

I explored objective, quantitative measures of aesthetic properties computed from relative distances between fiducial points. It is not possible to compare absolute distances since the magnification of the images was not standardized over the data set images. The pixel coordinate system was used since it preserves the unique coordinates for each fiducial point when the images are resized. I designed two measures relating to ptosis using the oblique and lateral clinical photographs.

Measure (2.1) is computed from the oblique views of the right and left breasts separately (using the proximal breast, which is nearest to observer's viewpoint). The vertical level of the sternal notch (s) is taken to be zero and the vertical displacements of the lateral terminus (i) and the nipple (n) are calculated from it. If the nipple is higher than the lateral terminus, then by definition there is no ptosis and the value is set to one; otherwise, measure (1) is calculated. Values near one indicate little ptosis while values near zero indicate significant ptosis (Figure 3.2).

$$\frac{s-i}{s-n} \tag{3.1}$$

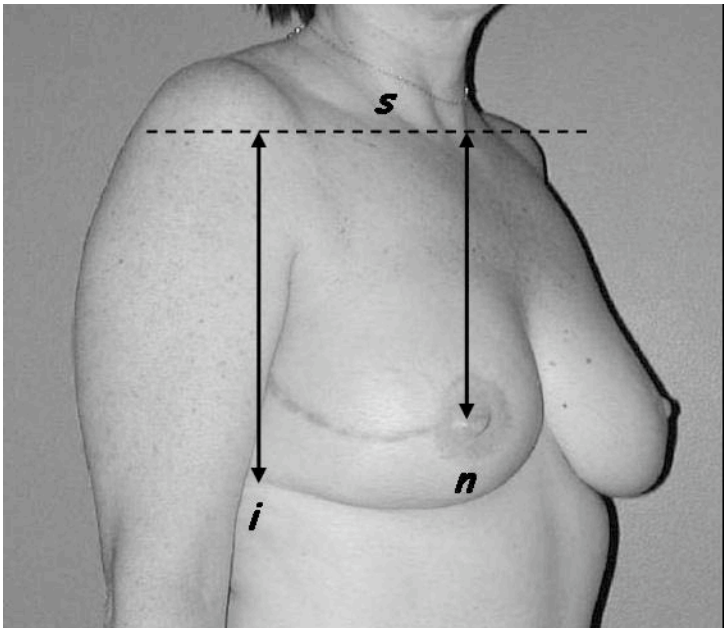


Figure 3.2: Measure (3.1) in right oblique view. The level of sternal notch (*s*) is extended with dotted line. The distances between sternal notch (*s*) to the point of lateral terminus of inframammary fold (*i*) and between the sternal notch (*s*) to the nipple (*n*) are indicated with arrowed lines.

Measure (3.2) is calculated from the lateral views of the right and left breasts separately. The vertical level of the lowest visible point of the breast (*v*) is taken to be zero, and the vertical displacements of the nipple (*n*) and the lateral terminus (*i*) are calculated from it. If the nipple is higher than the lateral terminus, then by definition there is no ptosis and the value is set to one; otherwise, measure (3.2) is calculated. Values near one indicate little ptosis and values near zero indicate significant ptosis (Figure 3.3).

$$\frac{n-v}{i-v} \tag{3.2}$$

Our objective measures for ptosis as defined above are expected to have values in the range from 0 to 1, but the observer rating scale ranges from 0 to 3. Thus, I rescaled

our objective measure values to be more similar to the observer rating scales to simplify interpretation of our objective measures. For the right and left breast separately of 10 cases, the observer ratings and objective measures were averaged across the three clinical observers (GPR, MJM, EKB) across the three time points. A simple linear regression was then used to relate the objective measures to the observer ratings. The regression model was used to transform the objective measures of the novice group for same 10 patient cases to evaluate the impact of the variability in manual identification of the fiducial points.

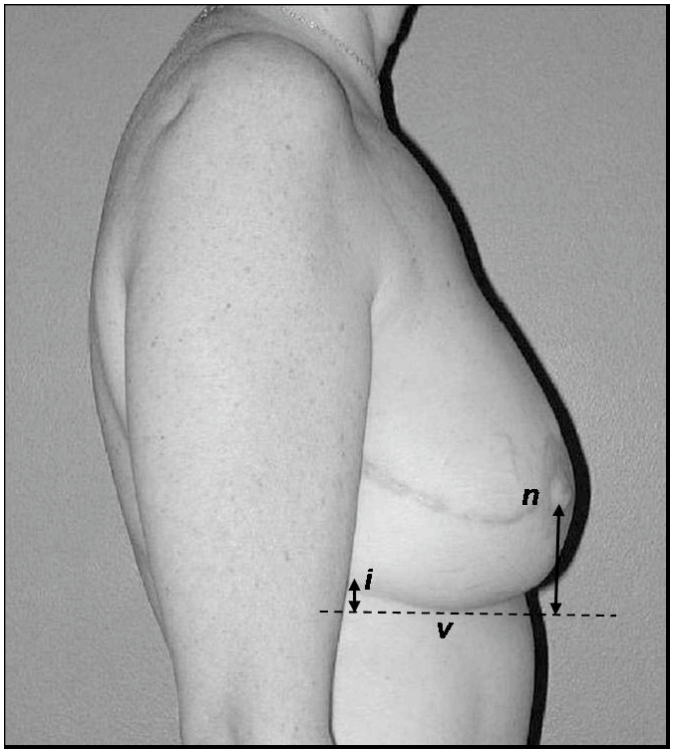


Figure 3.3: Measure (3.2) in right lateral view. The level of lowest visible point (*v*) is extended with dotted line. The distances between nipple (*n*) to the lowest visible point (*v*) and between the point of the lateral terminus of inframammary fold (*i*) to the lowest visible point (*v*) are indicated with arrowed lines.

3.3 RESULTS

3.3.1 Ptosis: observer rating measures

Kappa statistics of observer rating scale measures between time points (1st vs. 2nd, 2nd vs. 3rd, and 1st vs. 3rd) and between the three clinical observers (GPR vs. MJM, GPR vs. EKB, and MJM vs. EKB) were studied. Kappa results showed good to excellent intra-observer agreement (0.52 - 0.84). However, the inter-observer agreement varied across observers (0.23 - 0.49) and was lower in the case of MJM vs. EKB (0.07 - 0.15). The average values of the observer ratings across the three clinical observers for 10 cases were used to calibrate our objective, quantitative measures (Table 3.2 and Table 3.3).

Table 3.2: Kappa statistics of observer ratings over the time by three clinical experts. All 52 patients images were rated for ptosis by three clinical observers (GPR, MJM, & EKB) using a 4-point scale (Table 3.1) based on Regnault [16] and Bostwick [15] at three points, approximately 2 weeks apart. The amount of time between ratings was chosen to minimize the likelihood that the clinical observer would recall his previous ratings. First vs. second, first vs. third, and second vs. third ratings are compared.

	First vs. Second			First vs. Third			Second vs. Third		
	GPR	MJM	EKB	GPR	MJM	EKB	GPR	MJM	EKB
Right Breast	0.64	0.59	0.63	0.67	0.52	0.61	0.78	0.84	0.67
Left Breast	0.72	0.71	0.59	0.64	0.64	0.62	0.74	0.74	0.66

Table 3.3: Kappa statistics of observer ratings across the three clinical experts. All 52 patients images were rated for ptosis by three clinical observers (GPR, MJM, & EKB) using a 4-point scale (Table 3.1) based on Regnault [16] and Bostwick [15] at three points, approximately 2 weeks apart. The amount of time between ratings was chosen to minimize the likelihood that the clinical observer would recall his previous ratings. GPR vs. MJM, GPR vs. EKB, and MJM vs. EKB ratings are compared.

	First			Second			Third		
	GPR vs. MJM	GPR vs. EKB	MJM vs. EKB	GPR vs. MJM	GPR vs. EKB	MJM vs. EKB	GPR vs. MJM	GPR vs. EKB	MJM vs. EKB
Right Breast	0.39	0.37	0.08	0.49	0.27	0.12	0.49	0.35	0.15
Left Breast	0.26	0.37	0.08	0.39	0.23	0.08	0.34	0.28	0.07

3.3.2 Fiducial point localization

I studied the variability of the three clinical observers over time, variability between the clinical observer group and novice group, and the time required for each fiducial points between the clinical observer group and novice group.

Figure 3.4 illustrates examples of the variability of a clinical observer (GPR) over three time points and Figure 3.5 illustrates the variability of three clinical observers (GPR, MJM, EKB) for one time point (Table 3.4 and Table 3.5). The clinical observer group was especially consistent in marking the locations of the nipples relative to novice group. In the oblique view, the maximum standard deviation across time points among all patients, both breasts was 1.64 pixels in x and 0.71 pixels in y for the nipples. In the lateral view, the maximum standard deviation across time points among all patients, both breasts was 1.27 pixels in x and 1.41 pixels in y for the nipples. The variability across time points among all patients, both breasts was likewise very small for the y coordinate

of the lowest visible point in the lateral view (1.15 pixels). The variability was moderately small (standard deviation ≈ 1.5 pixels) for most of the other fiducial points, the exception being the y coordinate of the lateral terminus of the inframammary fold (9.12 pixels). Over all, the clinical observers were extremely consistent over time in marking all of the fiducial points used in this study.

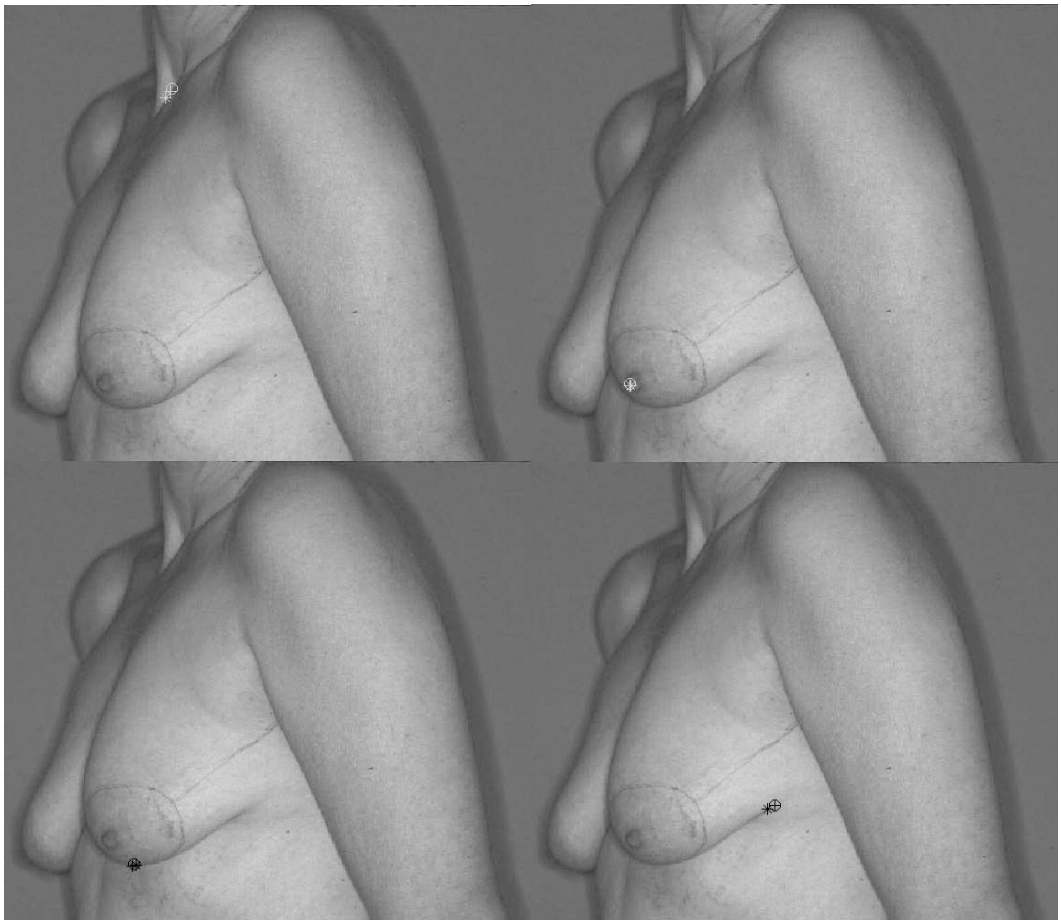


Figure 3.4: Variability of fiducial point localization by expert over time (+: 1st trial, o: 2nd trial, ✲: 3rd trial). Expert observer (GPR) marked the fiducial points (left top: sternal notch, right top: left nipple, left bottom: the lowest visible point, right bottom: lateral terminus of inframammary fold). The sternal notch and left nipple are marked with white markers and the lowest visible point and the lateral terminus of the inframammary fold are marked with black markers Four left oblique views of one patient were generated for example photographs.

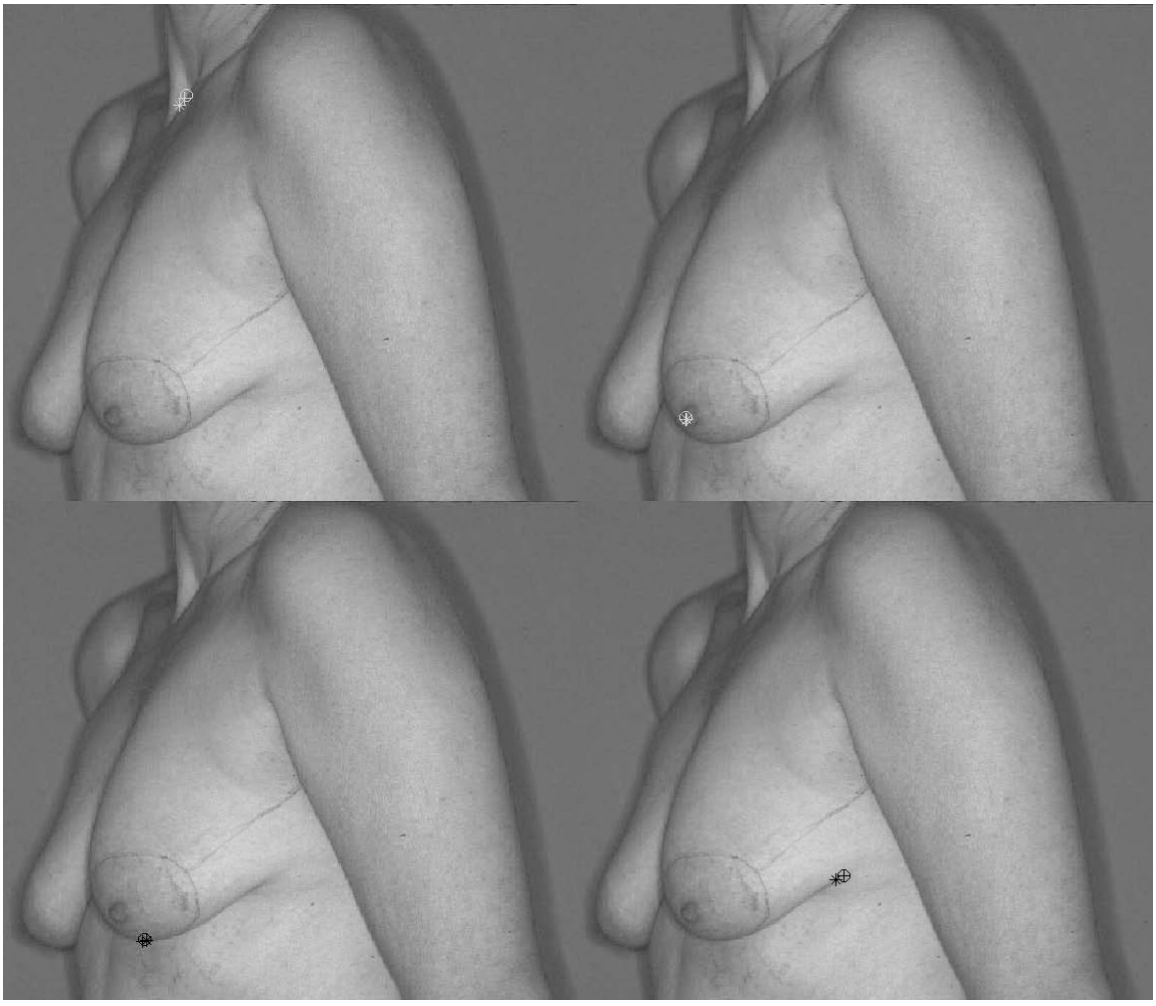


Figure 3.5: Variability of fiducial point localization by three experts at first trial (+: GPR, o: MJM, ✲: EKB). Three expert observers (GPR, MJM, EKB) marked the fiducial points (left top: sternal notch, right top: left nipple, left bottom: the lowest visible point, right bottom: lateral terminus of inframammary fold). The sternal notch and left nipple are marked with white markers and the lowest visible point and the lateral terminus of the inframammary fold are marked with black markers Four left oblique views of one patient were generated for example photographs.

Figure 3.6 and Figure 3.7 illustrate examples of the variability between a clinical observer (GPR) and non-clinical observers. As is shown in these examples, overall the novice observers in this study were able to accurately mark the locations on the fiducial

points if the clinical observer's marks are taken as the ground truth. Table 3.4 and Table 3.5 report the maximum standard deviation across time for each fiducial point among all novice observers viewing all patients, both breasts. As was the case for the clinical observer group, the novice group shows the least variability of all the fiducial points for localizing the nipples in either view (standard deviations ≈ 1 pixels). In fact, the variability in localizing most of the fiducial points was similar for the clinical and novice observers, with two exceptions. In the oblique view, the novice observers showed considerably more variability for the x- and y- coordinates of the lateral terminus of the inframammary fold and the y coordinate of the sternal notch. These structures are among the more subtle fiducial points used in this study and these results indicate that additional training may be desirable for novice observers or that automated methods should be developed.

Figure 3.6: Variability of fiducial point localization comparing an expert and novice groups for Measure (3.1). Observers marked the fiducial points (left: sternal notch, middle: left nipple, right: lateral terminus of inframammary fold). Fiducial points marked by six observers are indicated with markers; +: expert observer and five novice observers: o, ✻, □, ◊,△). The sternal notch and left nipple are marked in white and the lateral terminus of the inframammary fold is marked in black. Three left oblique views of one patient were generated for example photographs.

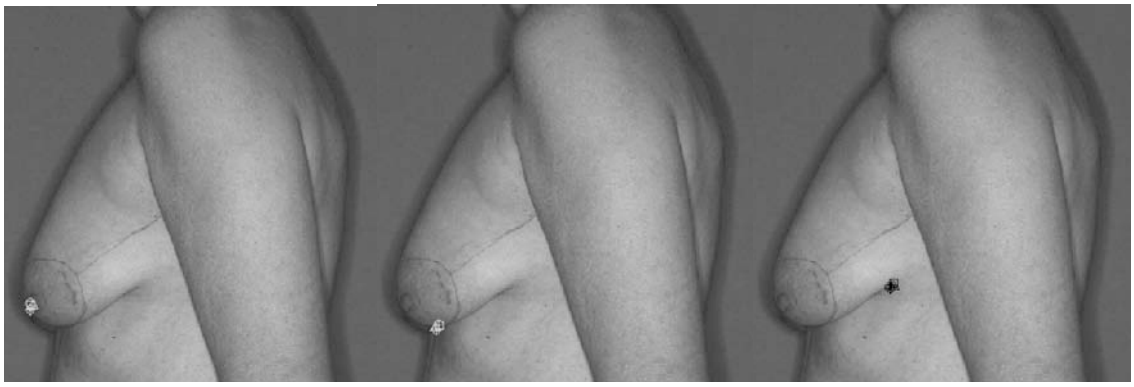


Figure 3.7: Variability of fiducial point localization comparing an expert and novice groups for Measure (3.2). Observers marked the fiducial points (left: left nipple, middle: lowest visible point, right: lateral terminus of inframammary fold). Fiducial points marked by six observers are indicated with markers; +: expert observer and five novice observers: o, ✻, □ , ◊,△). The left nipple is marked in white and the lowest visible point and the lateral terminus of the inframammary fold is marked in black. Three left oblique views of one patient were generated for example photographs.

Table 3.4: Variability of marking fiducial points localization between clinical and novice groups for 10 patients (Figure 3.6). Eight observers (three clinical and five novices) marked the fiducial points using our new objective tools. Standard deviations of the fiducial points from Measure (3.1) (right and left oblique views) are tabulated. Minimum and maximum values of standard deviations from x- and y- coordinate are calculated across the three time points for each patient and each breast.

(Clinical/ non-clinical) Unit: pixel coordinates	Minimum (x-coordinate)	Maximum (x coordinate)	Minimum (y-coordinate)	Maximum (y-coordinate)
Sternal notch *(s)*	0.43 / 1.22	4.59 / 5.26	0.26 / 0.80	3.97 / 8.54
Nipples *(n)*	0.25 / 0.18	1.64 / 1.79	0.25/ 0.41	0.71 / 2.74
Lateral terminus of Inframammary fold *(i)*	0.25 / 2.01	6.38 / 12.50	0.71 / 1.32	5.30 / 11.38

Table 3.5: Variability of marking fiducial points localization between clinical and novice groups for 10 patients (Figure 3.7). Eight observers (three clinical and five novices) marked the fiducial points using our new objective tools. Standard deviations of every fiducial points from Measure (3.2) (right and left lateral views) are tabulated. Minimum and maximum values of standard deviations from x- and y- coordinate are calculated across the three time points for each patient and each breast.

(Clinical/non-clinical) Unit: pixel coordinates	Minimum (x coordinate)	Maximum (x coordinate)	Minimum (y coordinate)	Maximum (y coordinate)
Nipples *(n)*	0.14 / 0.41	1.27 / 2.18	0.11 / 0.43	1.41 / 1.82
Lateral terminus of Inframammary fold *(i)*	0.56 / 1.95	6.36 / 16.18	0.73 / 1.38	9.12 / 11.35
Lowest visible Point *(v)*	0.25 / 0.47	3.28 / 2.87	1.01 / 0.28	2.31 / 1.46

The time required for marking fiducial points by both the clinical and novice groups for 10 patients was studied (Table 3.6). The mean time required to locate each fiducial point in each trial was computed. The mean times required for marking all fiducial points for Measure (3.1) and Measure (3.2) were 46.44 seconds and 46.01 seconds for the clinical and novice groups respectively. This result indicates that non-surgeons can identify the fiducial points needed for our objective measures as quickly as surgeons can. It was also observed that the time needed to locate the fiducial points typically decreased with each trial for both the clinical and novice groups, presumably as a result of the users becoming more accustomed to the software interface.

Table3.6: Time required for marking fiducial points between clinical and novice groups for 10 patients. Eight observers (three clinical and five novices) marked the fiducial points using our new objective tools. Mean values of time for each fiducial point and each trial from Measure (3.1) and Measure (3.2) are computed.

	Clinical Group 1st / 2nd / 3rd / overall	Novice Group 1st / 2nd / 3rd / overall
Nipples (n)	4.56 / 3.84 / 3.56 / 3.99	3.75 / 3.11 / 2.95 / 3.27
Lateral terminus of Inframammary fold (i)	4.27 / 3.43 / 3.01 / 3.51	5.35 / 3.95 / 3.35 / 4.22
Lowest visible Point (v)	3.73 / 3.45 / 3.35 / 3.51	4.20 / 3.85 / 3.15 / 3.74
Sternal notch (s)	5.34 / 4.12 / 4.33 / 4.59	5.25 / 3.99 / 3.65 / 4.30
Time required for all fiducial points in Measure (3.1) and Measure (3.2) per patient (four nipple points, four lateral terminus of Inframammary folds, two lowest visible points, two sternal notches)	46.44	46.01

3.3.3 Ptosis: objective measures

Our objective measures for ptosis are expected to have values in the range from 0 to 1, but the observer rating scale ranges from 0 to 3. I rescaled the values of our new measures to be more similar to the observer rating scales to simplify interpretation of our objective measures. The mean of the three clinical observers' observer ratings at three time points were calculated over right and left breasts separately for 10 patient cases. Simple linear regression was used to relate the ptosis ratings obtained using the observer rating scale to our objective measures (Figure 3.8, Figure 3.9).

The simple linear regression model from clinical group was used to transform the objective measures from novice group's distance ratios to a scale more similar to that of the observer rating scale to allow for easier interpretation of the objective measures. After rescaling the objective measures, I determined that the level of variability of the novice group in the objective ptosis measure resulted from variability in the fiducial point marking. Maximum variability from different observers for the same case at different time point was about 0.47 ptosis grade on Measure (3.1) (Table 3.7) and 0.42 ptosis grade on Measure (3.2) (Table 3.8).

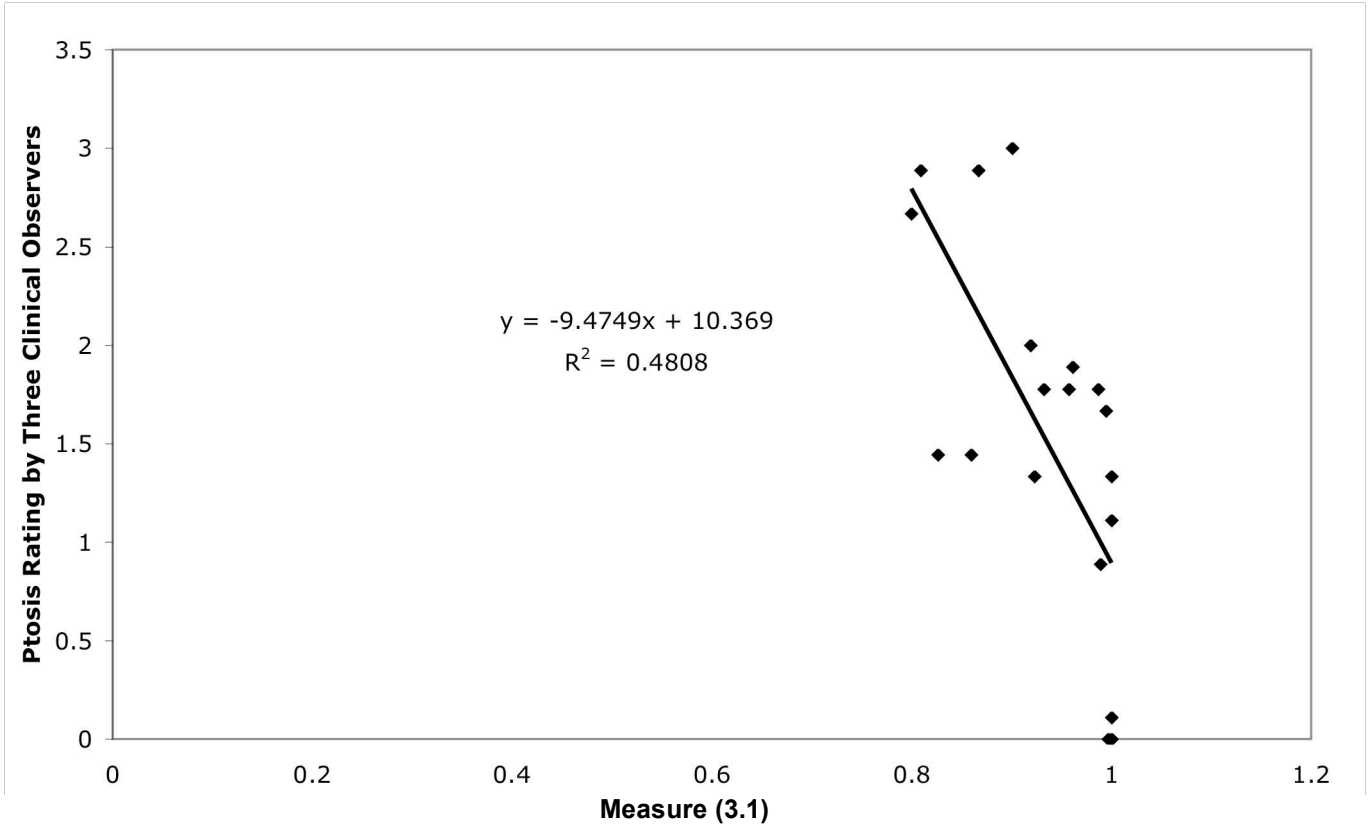


Figure 3.8: Simple Linear Regression between ptosis ratings made by the clinical observer and Measure (3.1). A simple linear regression was used to relate average values computed for the objective measures to the observer ratings for 10 patient cases provided by three clinical observers (GPR, MJM, EKB). The regression data was used to transform the objective measure data of novices group for same patient cases to evaluate the impact of the variability from manual identification of the fiducial points.

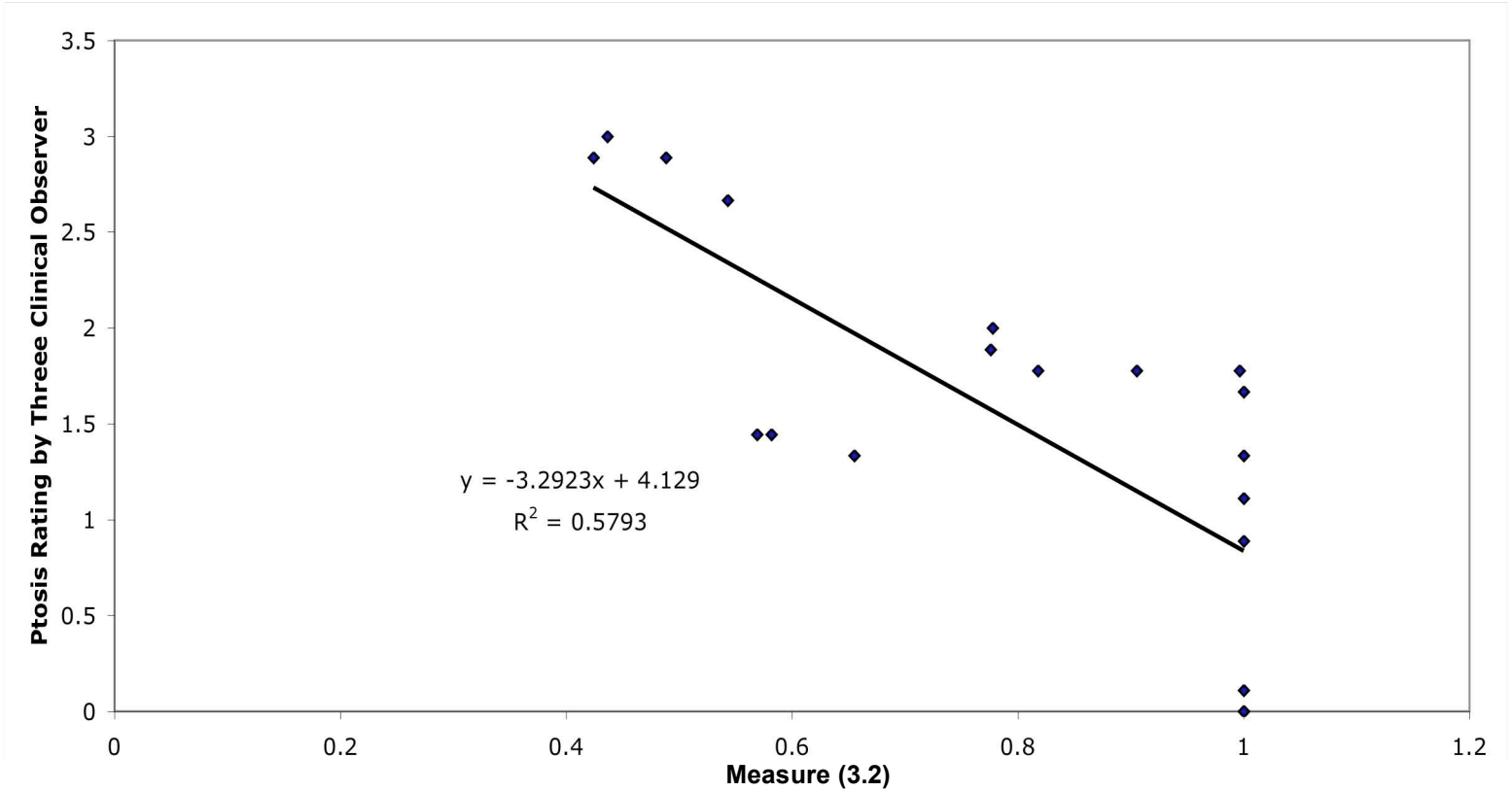


Figure 3.9: Simple Linear Regression between ptosis ratings made by the clinical observer and Measure (3.2). A simple linear regression was used to relate average values computed for the objective measures to the observer ratings for 10 patient cases provided by three clinical observers (GPR, MJM, EKB). The regression data was used to transform the objective measure data of novices group for same patient cases to evaluate the impact of the variability from manual identification of the fiducial points.

Table 3.7: Maximum variability of ptosis grade after rescaling all five non-clinical observers' objective Measure (3.1). Standard deviations of right and left breast are tabulated separately.

	Right Breast	Left Breast
Patient #1	0.27	0.37
Patient #2	0.11	0.31
Patient #3	0.14	0.39
Patient #4	0.10	0.33
Patient #5	0.24	0.30
Patient #6	0.18	0.12
Patient #7	0.47	0.27
Patient #8	0.09	0.35
Patient #9	0.09	0.19
Patient #10	0.06	0.26

Table 3.8: Maximum variability of ptosis grade after rescaling all five non-clinical observers' objective measure (3.2). Standard deviations of right and left breast are tabulated separately.

	Right Breast	Left Breast
Patient #1	0.37	0.21
Patient #2	0.17	0.24
Patient #3	0.24	0.30
Patient #4	0.24	0.21
Patient #5	0.21	0.22
Patient #6	0.42	0.18
Patient #7	0.27	0.24
Patient #8	0.07	0.27
Patient #9	0.00	0.16
Patient #10	0.02	0.18

Two additional simple linear regression analyses were performed for comparison, which I will refer to as models B and C. Regression model B was built on all 52 cases using the observer ratings from clinical group vs. the objective measurements based on the fiducial point markings of a single novice observer (MSK). The same modeling strategy was used for model C, with the difference that the 10 patient cases which were used for testing the rescaling of the novice group measurements were excluded. The results indicated that the maximum variability from different novice observers from model B were 0.50 ptosis grade on Measure (3.1) and 0.46 grade on Measure (3.2) and for model C were 0.48 ptosis grade on Measure (3.1) and 1.25 ptosis grade on Measure (3.2). Thus, the results of using model B, which used a larger number of cases but a smaller number of observers than the main model, were essentially equivalent to that of the main model. The results of using model C, in which the 10 cases being assessed were held out when the model was built, showed an increased variability as was expected.

3.4 DISCUSSION AND CONCLUSION

In this study, I designed quantitative, objective measurements of breast ptosis based on ratios of distances between fiducial points manually identified in lateral and oblique views of clinical photographs. Three clinical observers rated preoperative clinical photographs of 52 patients who underwent breast reconstruction for ptosis using a qualitative, observer rating scales at three time points (at least 2 weeks). Kappa statistics between time points indicated good to excellent intra-observer agreement (0.52-0.84), which encouraged our use of the average of the three observers' ratings as a guide in assessing our new quantitative, objective measurements of breast ptosis.

In our current implementation, fiducial points are manually identified. While still not ideal, it would be more practical to use manually identified fiducial points if the task

could be accurately performed by a non-clinical observer (*e.g.*, a research assistant). In this study I found that novice observers could reliably locate the required fiducial points as compared to the identifications made by clinical observers. Moreover, both novices and clinical observers demonstrated low variability in repeated marking of fiducial points at different time points. Thus, it is practical to use our approach with novice observers. However, a more automated method would be better for large outcome studies or for developing intra-operative tools. It takes about 40 seconds per patient to mark the 12 fiducial points required for the two ptosis measures presented here, though additional fiducial points would likely be needed to address other aesthetic properties (*e.g.*, symmetry).

Our quantitative, objective measurements of breast ptosis are based on ratios of distances between fiducial points. Typically, only the y coordinate, but not both are needed for any particular fiducial point. For measure (3.2), it was encouraging that for the lateral terminus of the inframmamary fold the coordinate used (y) was less variable than the one that wasn't needed from the lateral view (x). A similar, but weaker trend was seen for the inframmamary fold in the oblique view used for measure (3.1). In future studies, the effect of the patient's body-mass index should be examined since the small sample used in this study suggests that some landmarks such as the sternal notch are more difficult to locate on heavier patients.

A simple linear regression model was developed to relate the observer ratings and objective measurements for 10 patient cases by the clinical observers. The regression model was used to rescale the objective measurements of the novice group to allow for easier interpretation of the variability in the measures due to variability in fiducial point identification. The objective measures from novice group showed that the maximum variability from different novice observers for the same cases at different time point was

about 0.47 ptosis grade on Measure (3.1) (Table 3.7) and 0.42 ptosis grade on Measure (3.2) (Table 3.8).

Over all there was good agreement between the novice and clinical observers, but some differences were seen. While both clinical and novice observers showed the same trend of some fiducial points being easier to reliably locate than others (*e.g.*, lateral terminus of inframammary fold *vs.* nipples), the differences between the easier and more difficult points were larger for the novice observers. These observations suggest that additional training may be desirable for novice observers, particularly for more subtle structures, such as the sternal notch.

One potential limitation of objective measurements of breast ptosis based on ratios of distances between fiducial points is that they could be difficult to interpret. We addressed this problem by building a linear regression model that relates the objective measurements to observer ratings. The linear regression model was then used to rescale each ratio to provide a final quantitative, objective measure that can be interpreted in the same manner as the more familiar observer rating scale for ptosis. The level of variability of the objective measures of novice observers after rescaling was found to be equivalent to less than half of a point on the observer ptosis scale. This result indicates our new objective measures show a high level of correspondence with observer ratings of clinical observers.

Chapter 4: ASSESSMENT OF ARTIFICIAL SURGICAL SCARS: COMPARISON OF PHOTOGRAPHY AND COLORIMETRY

Contribution and publication: A preliminary version of the study in this chapter will be presented at the American Medical Informatics Association Annual Symposium and a manuscript is being prepared for submission to a journal.
1. M. S. Kim, W. N. Rodney, G. P. Reece, T. Cooper, and M. K. Markey, "Toward Quantifying the Aesthetic Outcomes of Breast Cancer Treatment: Comparison of Clinical Photography and Colorimetry," in *American Medical Informatics Association Annual Symposium* Chicago, Illinois, 2007.

4.1 INTRODUCTION

4.1.1 Breast surgical assessment

All types of breast surgery, including breast reconstruction, leaves scars that can be disfiguring, aesthetically unpleasant, and can negatively affect a patient's quality of life [151-153]. Scars can cause itching, tendernesits, pain, sleep disturbances, anxiety, depression, and disruption of daily activities. Other psychosocial complications include development of post-traumatic stress reactions, loss of self-esteem, and stigmatization [151, 152].

Scar tissue is physically different from normal tissue and can have abnormal color, rougher texture, and increased thickness and firmness [154]. The five main categories of scarring are defined by the growth and characteristics of the scar tissue [151]. Widespread scars appear when the fine lines of surgical scars gradually become stretched and widened, usually within three weeks of surgery. Atrophic scars are generally small and often round with an indented center below the surrounding skin. Scar contractures are scars that cross-joints or skin creases at right angles and are prone to develop shortening or contracture. Hypertrophic scars are raised scars that remain within the boundaries of the original lesion and typically occur after burn injury on the trunk and

extremities. Keloid scars are raised scars that spread beyond the margins of the original wound and invade the surrounding normal skin in a way that is site specific.

Assessment of aesthetic outcomes, including scarring, is important to protect patients' health and well-being [59], but there is no accepted method for assessing the appearance of surgical scars. Most assessments of surgical scars by clinical observers [94], patients [88], or combinations of both [91] have only been studied as subscales for aesthetic outcome measures (Table 4.1). Visibility, the extent the surgical scar detracts from the aesthetic outcome, is the accepted method to assess the scarring in breast reconstructive surgery.

Commonly used visual assessments are a 3-point scale (Good, fair, poor) [94, 155], a 4-point scale (Excellent, good, fair, poor) [95] or a 5-point scale (Excellent, good, satisfactory, poor, unacceptable) on breast mound scars. These global assessments have been based on postoperative photographs [153] or direct views of patients and show unacceptable inter-rater reliability [94] and inability to consider all of the relevant aesthetic features of scars.

Quantitative assessment methods study the mechanical properties, size, and color of the scar area (Table 4.2). However, while mechanical properties of scars influence the comfort of a patient, they do not impact the visibility, and thus the aesthetics, of scars. The visibility of a scar is significantly influenced by surface area, vascularity, and pigmentation [156].

Table 4.1: Summary findings of Observer Scar Assessments

Reference	Scale	Test Population	Findings
Veiga [155]	3-point scale with subscales (breast scars)	20 patients of unilateral mastectomy	The inter-rater and intra-rater agreement was poor to fair for the majority of subscales.
Lowery [94]	4-point, subscales (breast scars)	50 photographs of BRC patients	Suggested explicit criteria and to separate various components of the aesthetic result to improve the reliability of the assessment ($\kappa appa = 0.31$)
Andrade [88]	5-point, subscales (Appearance of surgical scars, Aesthetic result of breast reconstruction)	214 questionnaires returned by patients of BRC	Significant association was found between overall cosmetics and scars.
Edsander-Nord [91]	7-point, subscales (scar on the breast)	Photographs of 53 patients of TRAM-pedicled flap reconstruction	A strong correlation between the patients' and the panel's evaluations of the aesthetic outcome was seen; generally, the panel's evaluation of the aesthetic result of the breast correlated with the satisfaction of the patients.
Sneeuw [95]	4-point scale with subscales (scar, size, shape, color, firmness)	76 Photograph of BCT patients	Reported higher intra-rater agreement between the nurse and the oncologist ($\kappa appa = 0.64$) but lower inter-rater agreements between ratings by patients and clinical observers ($\kappa appa < 0.10$).
Godwin [153]	5-point scale, subscales (breast scar)	34 photograph of reduction mammaplasty	Scarring was the most frequent cause of dissatisfaction for both surgeons and patients.

Table 4.2: Summary findings of Quantitative Scar Assessments and Measures

Reference	Test Population	Findings
Truong [157]	59 breast cancer patients	VSS, patient self-rating scale, and Short-Form McGill Pain Questionnaire had acceptable internal consistency (Cronbach's α = 0.79, 0.64, 0.72). Patient satisfaction was not significantly associated with VSS scores.
Kar [156]	100 linear surgical scars	Internal consistency of observer and patient scales was good (Cronbach's α = 0.86 and 0.90, respectively) and reliability of the observer scale was good for the total score (*Pearson* $r = 0.96$, $p < 0.001$).
Sullivan [158]	73 burn patients	VSS is a useful tool for objective comparison of the same scar by different observers.
Draaijers [159]	49 scar areas in 20 patients	Reliable measurements with the DermaSpectrometer and Minolta Chroma-Meter (*Pearson* r = 0.72). At least three observers necessary with for reliable pigmentation measurement (*Pearson* $r >$ or = 0.77). Agreement for observers for pigmentation was unacceptable (*kappa* = 0.349).
Setaro [160]	348 photographs of Caucasian	Clinically useful digital measurement of absent, slight, and moderate erythema using normalized r, g, b color coordinate system.
Coelho [161]	Photographs of 9 subjects of White, Asian, or African-American	Feasible to use digital photography for objective evaluation of UV erythema and pigment changes due to UV exposure in different racial/ethnic groups even with light to heavy UV exposure schedules (*Pearson* r = 0.66, 0.96, 0.93).
Zuijlen [162]	20 normal volunteers	Planimetry by photography is more reliable than planimetry by tracings for areas that are not extremely curved.
Hurtut [163]	undefined number of scoliosis patient images	Encouraging results with an average deviation using automatic contours compared to manual contours of $0.67 \pm 0.24\, mm$ with a camera resolution of $1.8\ mm$. This is less than the deviation between two manual contours ($0.75 \pm 0.21\ mm$).

Since the size and color of scars have the largest influences on the overall visibility, quantitative ways of obtaining these data have been of great interest [156, 159, 164, 165]. Vascularity and pigmentation can be separately evaluated by observing the color of the scar before and after it is blanched using a tool such as a piece of clear plastic [156, 165]. However, for purposes of quantifying the visibility of scars, it is not necessary to distinguish between coloration due to vascularity vs. pigmentation [166]. Currently, observers assess the outcome of scars either by manual inspection, clinical instruments or combinations. This can be time consuming, demanding of the patient, and expensive with equipment costs. An effective, practical means of quantitatively assessing scar color is needed and digital images can be used for this assessment.

Towards this aim, this chapter will discuss the potential color assessment method on surgical scars using digital photograph and the following chapter will discuss area measurement methods on real scars.

4.1.2 Color assessment

To standardize the quantification of color, CIE (International Commission on Illumination) recommended tri-stimulus color values XYZ, which are read using spectrometric reflectance data from an object and three primary stimuli strictly defined by the commission. CIE L*a*b* (CIELAB) is the most complete color model used conventionally to describe all the colors visible to the human eye. The three parameters in the model represent the lightness of the color (L*, L* = 0 indicates black and L* = 100 indicates white), its position between magenta and green (a*, a* = -60 indicates green and a* = 60 indicates magenta), and its position between yellow and blue (b*, b* = -60 indicates blue and b* = 60 indicates yellow) [167].

While for purposes of assessing scar visibility it is not necessary to separate the color contributions from the underlying physiological factors of vascularity and pigmentation, for completeness I briefly review how the color parameters are known to relate to the skin physiology. Studies have shown that L* and melanin index correlate almost linearly if the Hb (oxyhemoglobin) amount is constant. The a* and erythema index values correlate almost linearly with the amount of Hb in the superficial plexus.

These findings suggest that L*, b*, or combinations of them are reasonable parameters for evaluation of the degree of pigmentation if the cutaneous blood volume is considered to be similar at all sites. The a* index is a good parameter for evaluating the degree of erythema or cutaneous blood volume. The difference values between the test and control sites are the best parameters for evaluating erythema and pigmentation [168]. These findings suggest that a combination of L* and b* can be used for measuring pigmentation and a* can be used for evaluating the redness of surgical scars. Although all of the L*a*b* parameters are important indicators of vascular changes or pigmentation, the a* parameter is the most relevant and sensitive one for erythemous or vasodiatatory change [167].

The drawback to the use of colorimeters is that they require direct patient interaction and can cost thousands of dollars. By comparison, digital photography is already in widespread use to document clinical outcomes. Moreover, digital photographs can be evaluated after the fact and so required less patient interaction. Thus, assessment of scar coloration by digital photography is an attractive alternative.

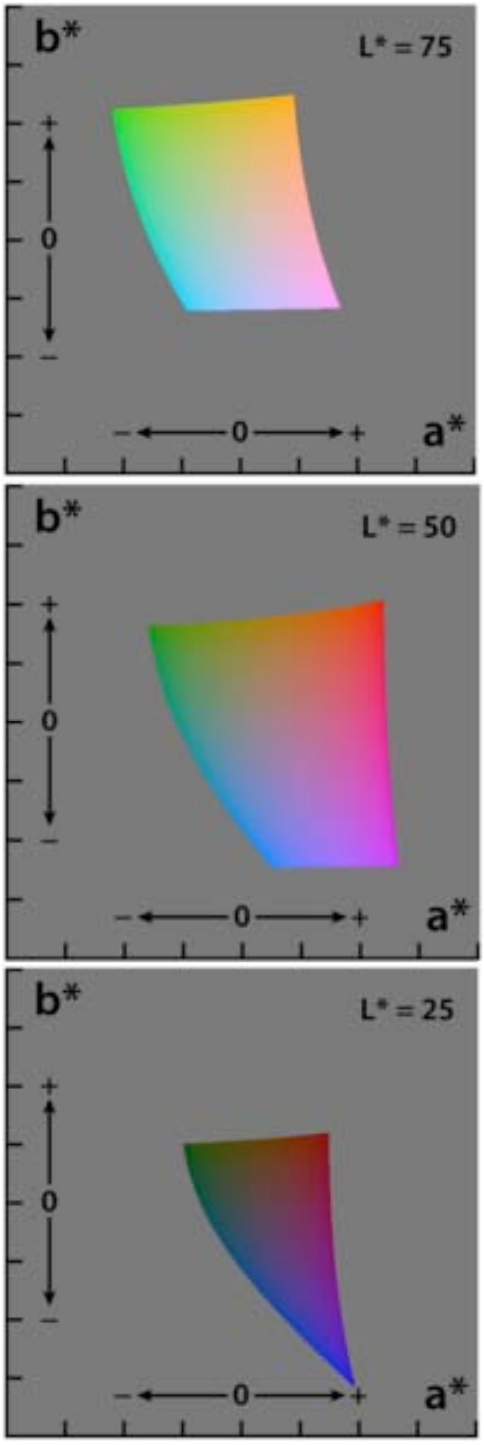


Figure 4.1: The L*a*b* color space, showing only colors that fit within the sRGB gamut (and can therefore be displayed on a typical computer display). Each axis of each square ranges from -128 to 128. (Reprinted, by permission granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".)

Towards this aim, it is encouraging that digital photography has been demonstrated to be a viable alternative to color measurements by a colorimeter for objective evaluation of UV erythema. Coelho *et al*. used the change in the a* parameter to quantify erythema, and ΔL was used to evaluate the level of pigmentation that develops during tanning for three subjects in each of three ethnic/racial groups: Asian, Black or African American, and White [169]. They reported their computer assisted digital image evaluation (CADIE) system, which uses a commercial digital SLR camera, a Nikon D1 (Nikon, Tokyo, Japan), and image processing software, Adobe Photoshop 6.0 (Adobe Systems, San Jose, CA), produced results comparable to those obtained with a spectroscopy system, the Minolta CM 2002 (Minolta, Ramsey, New Jersey). They found good agreement between the Δa* (Pearson r = 0.77-0.97) and ΔL* (Pearson r = 0.66-0.96) generated by the two methods.

Careful experimental control and standardization is needed in order to collect precise color information. Past studies found that an artificial skin model could be successfully created to control the experimental environment and to demonstrate the diversity of color of skin regions [170, 171]. In particular, a multilayered skin model was developed to measure the amount of melanin and hemoglobin with a microscope interfaced with a computer and the values were compared with those of real skin regions. Similarly, in this study, artificial scars created by a makeup artist were employed in order to achieve maximal control of the experimental set up, i.e., there was no hemoglobin in the makeup used to create the artificial scar.

In this chapter, I compared color measurements obtained by clinical digital photography to those from a standard colorimeter. Experimental conditions were controlled by performing measurements on artificial scars created by a makeup artist (TC). The colorimeter measurements of the artificial scars were compared to those

reported in the literature for real scars in order to confirm the validity of this approach. Agreement in photographic and colorimeter measurements was statistically analyzed using the hypothesis test for equivalence [172], the ICC coefficient [173], and the Bland-Altman method [174, 175].

4.2 MATERIALS AND METHODS

4.2.1 Artificial scar creation on models

Two female models in their early twenties representing two major race groups (Model A: White and not Latino or Hispanic, Model B: Black or African American and not Latino or Hispanic) were hired for the study. Their demographic and clinical background information is summarized in Table 4.3. Neither of the two models has a history of pregnancy or lactation, but Model B has a history of breast surgery on her left side approximately 7 years ago.

A makeup artist (TC) created 6 types of scars while taking into consideration the reality of the scar appearance from a distance, excessive shine, and the utilization of latex or collodion makeup (Table 4.4). An experienced plastic surgeon (GPR) guided the creation of artificial scars that emulate the appearance of scars that would result after breast reconstruction with a latissmus dorsi (LD) flap or a transverse rectus abdominis myocutaneous (TRAM) flap. In LD reconstruction, the flap is composed of an elliptical patch of skin that lies over the latissimus dorsi muscle. Typically, the flap is used to replace the breast skin removed during the mastectomy and the breast volume is replaced with a breast implant [65]. The TRAM is the most popular form of tissue-based breast reconstruction, wherein a flap from the abdomen, either with blood vessels intact or completely severed from the body, is transferred to the chest either via a subcutaneous

tunnel or using microsurgical technique to replace the breast skin and volume [70, 71, 176-178]. Note that artificial donor site scars were not created in this study.

Table 4.3: Clinical and demographic background information of Model A and Model B. None of two models had history of pregnancy or lactation, but Model B has a history of breast surgery on left side about 7 yeas ago, which might affect the aesthetic assessment of the breast.

	Model A	Model B
Age (in years)	21	22
Weight (in kg)	57	64
Height (in cm)	170	173
Bra size	36	36
Cup size (A, B, C, D)	B	C
Ethnicity	Not Hispanic or Latino	Not Hispanic or Latino
Race	White	Black or African American
Tanning booth exposure	Have had blister burns. Not more than 4 hours of sun per week	Never tan in a booth Sometimes sunbath outside no more than an hour, with sun block though. SPF 50.
Sunscreen use	No routine use of sunscreen	Always. And waterproof.
Pregnancy and lactation history	None	None
Estrogen use	Oral contraceptive	Oral contraceptive
Previous breast surgery	None	Had a benign tumor in left breast removed at 15
Fitzpatrick Skin Type (FST) [179]	II. Burn and minimal tan, Pale white	VI, Tan, no burn, Dark brown
Notes on FST	Mainly burns, but little tan	Have never had a sun burn
Ptosis grade [5, 180]	0: Nipple and most of gland are above IMF	2: Nipple is below IMF, but above lower contour of the breast
Pseudoptosis	No	No

Table 4.4: Artificial scars created on Models A and B. The scars created for Model A resemble normal, hypertrophic scars that are raised scars that remain within the boundaries of the original lesion typically seen 3 to 6 months (Figure 4.2, Figure 4.3, Figure 4.4). The scars created for Model B resemble normal, hypertrophic scars (Figure 4.5, Figure 4.6) and keloid scar (Figure 4.7), typically seen 6 months to 1 year after surgery. All scars were created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area.

	Scar A1	Scar A2	Scar A3	Scar B1	Scar B2	Scar B3
Location						
Side	Left	Left	Left	Left	Left	Left
O'clock	3	3	3	4	4	2
Makeup Base	Rigid collodion	Rigid collodion	Rigid collodion	Rigid collodion	Rigid collodion	Rigid collodion
Height (in mm)	0	None	1	0	0	4
Width (in mm)	3	None	3	2	6	4
Length (in mm)	5	None	5	6	9	9

The scars created for Model A resemble normal scars (Figure 4.2, Figure 4.3), and hypertrophic scars (Figure 4.4), raised scars that remain within the boundaries of the original lesion, as typically seen 3 to 6 months after surgery. Scar A1 was created on left breast around the nipple-areola complex (NAC) with the incision running into the low superior lateral area. Scar A2 was started from Scar A1 and the scar was expanded around the NAC into the medial region of the left breast. Liquid latex was added on Scar A2 to make the nipple more asymmetric in photographs of Scar A3.

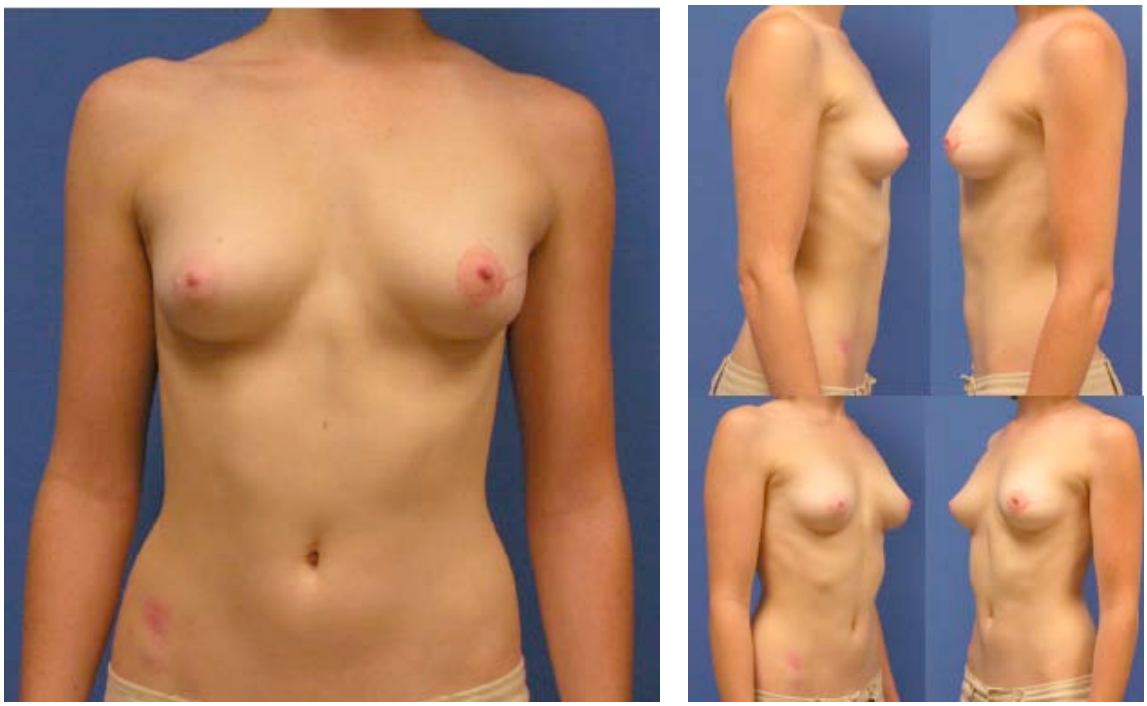


Figure 4.2: 2D image of Scar A1 on Model A. The scars created for Model A resemble normal scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scar A1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

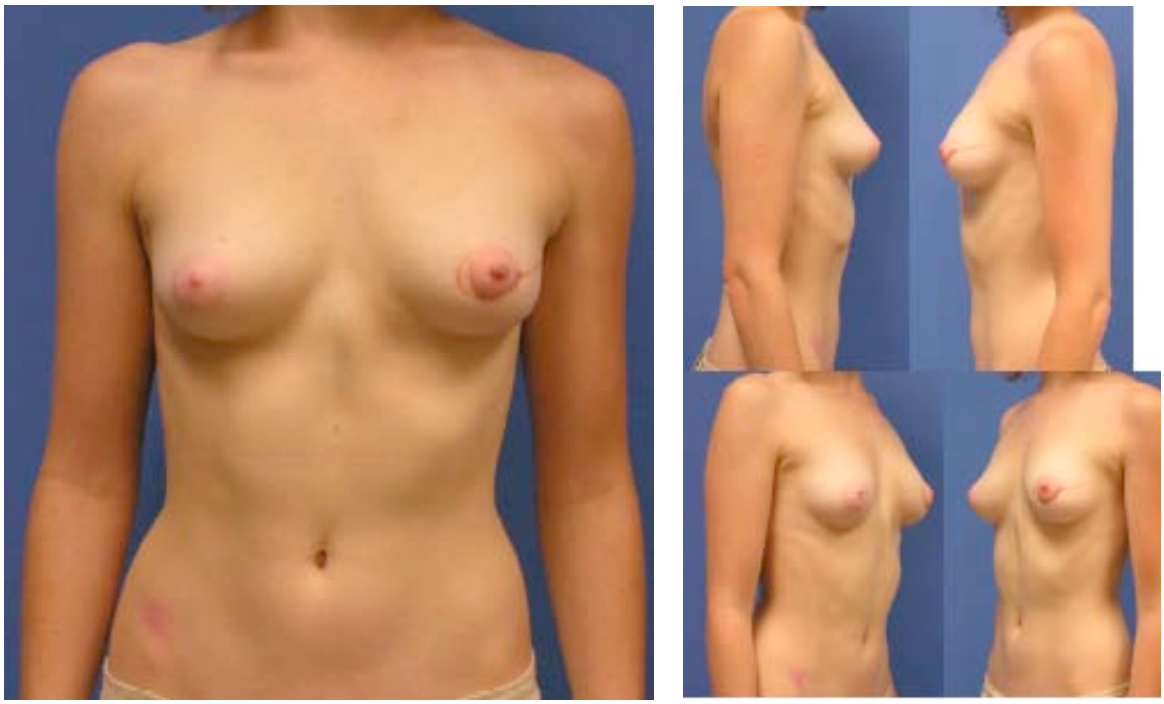


Figure 4.3: 2D image of scar A2 on Model A. The scars created for Model A resemble normal scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. The scar was expanded around the nipple area into the medial region of the breast. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

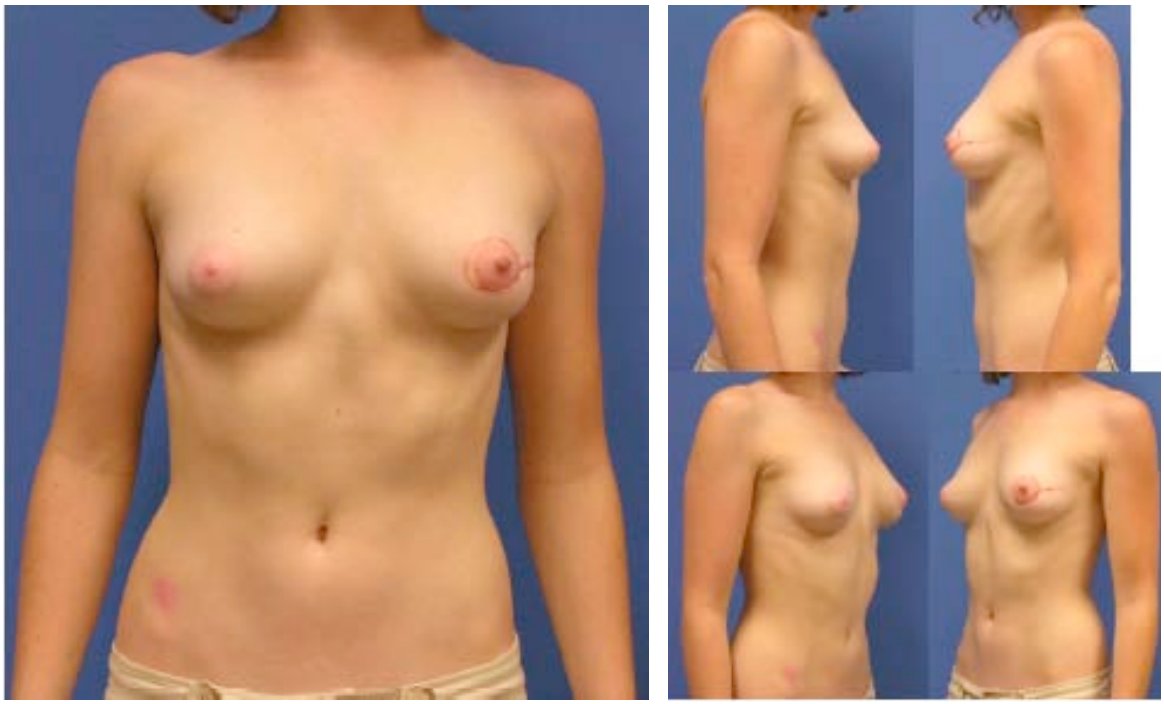


Figure 4.4: 2D image of scar A3 on Model A. The scars created for Model A resemble hypertrophic scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Liquid latex was added to make nipple more asymmetric on Scar A3.

The scars created for Model B resemble normal scars (Figure 4.5), hypertrophic scars (Figure 4.6), and a keloid scar (Figure 4.7), a scar that has spread beyond the margins of the original wound and invades the surrounding normal skin [59]. The scars on Model B resemble those typically seen 6 months to 1 year after surgery. Scar B1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scar B2 was started from Scar B1 and crescent shaped region to lateral side of areola was added. Scar B3 was built off from Scar B2 and added rigid collodion to create a keloid appearance.

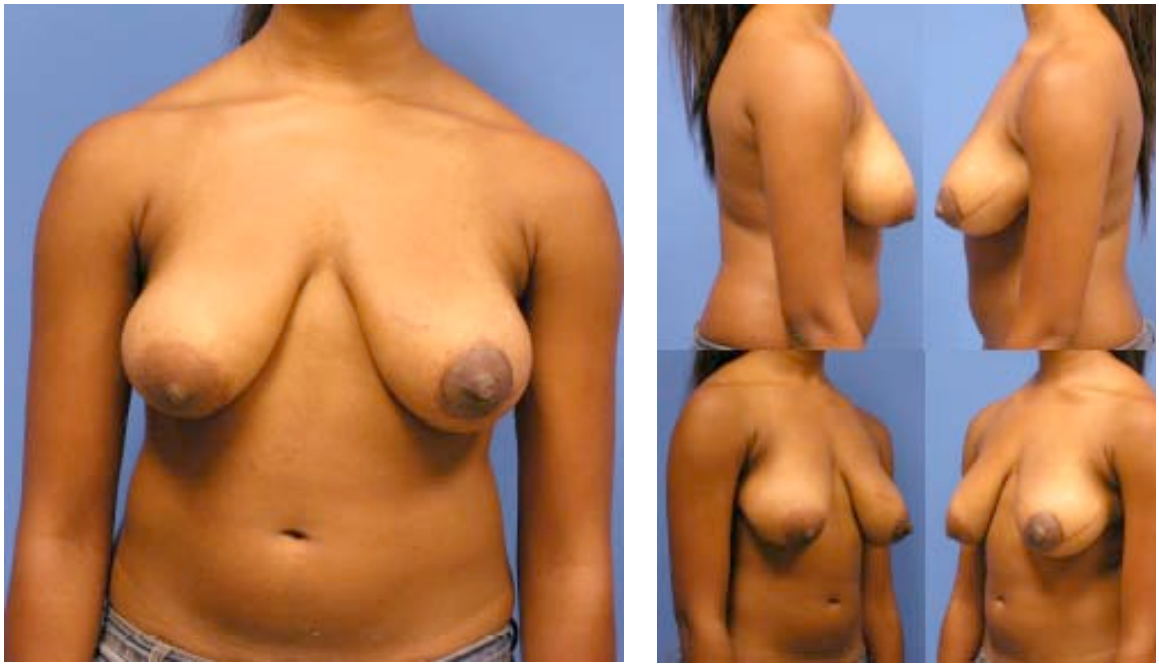


Figure 4.5: 2D image of scar B1 on Model B. The scars created for Model B resemble normal scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scar B1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

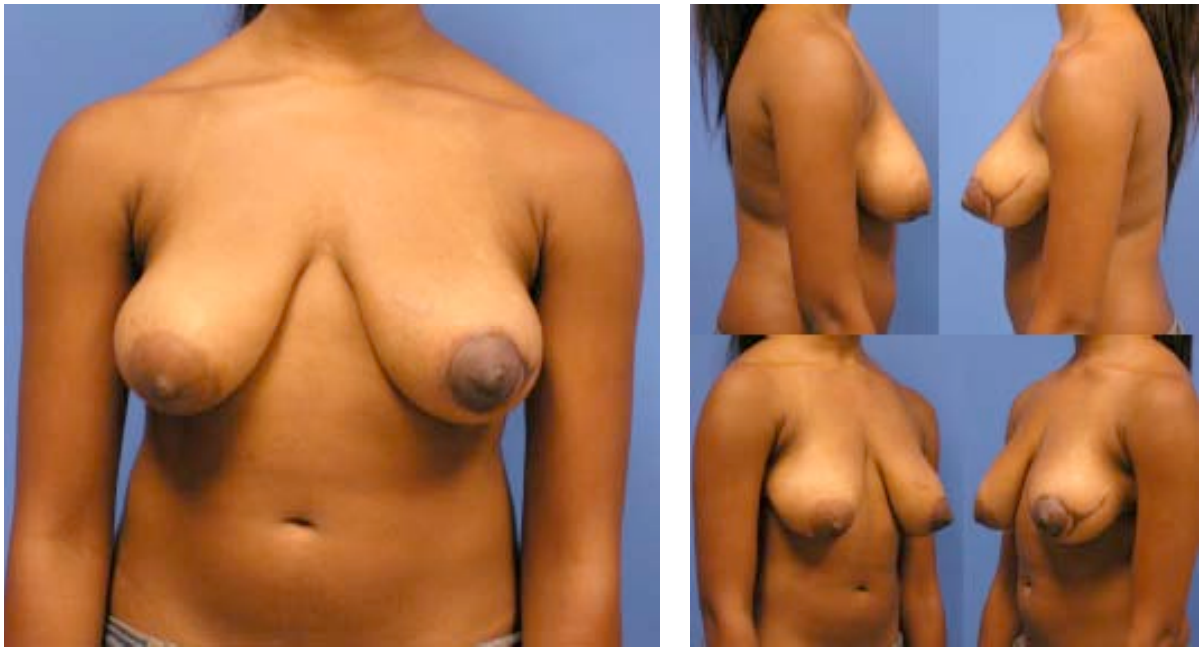


Figure 4.6: 2D image of scar B2 on Model B. The scars created for Model B resemble hypertrophic scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Scar B2 was started from Scar B1 and crescent shaped region to lateral side of areola was added.

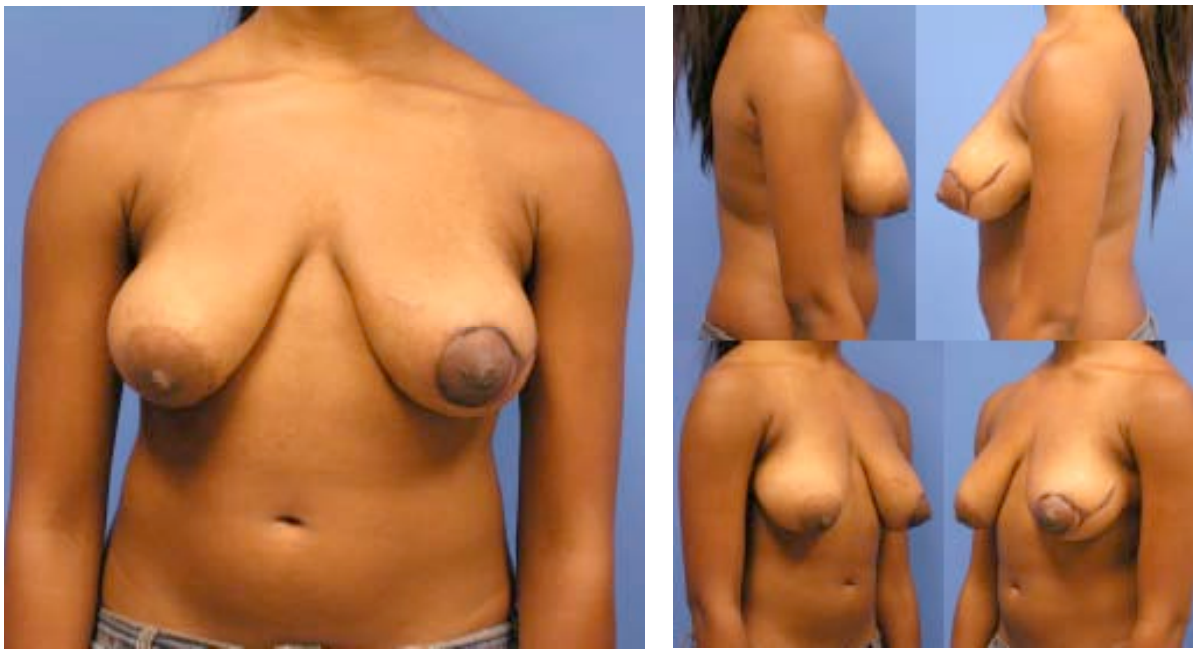


Figure 4.7: 2D image of scar B3 on Model B. The scars created for Model B resemble keloid scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Scar B3 was built off from Scar B2 and added rigid collodion to create a keloid appearance

The basic scar shape was initially created using liquid latex, rigid collodion, or a combination of these products on top of alcohol cleansed skin in order to get the desired effect. The simulated scars were then covered with a natural skin-toned base foundation to help blend the edges of the created scars with the bare skin. Top detailing color makeup in various tones was next meticulously placed over the simulated scars to give them a more organic appearance. Neutral set powder was pressed over the simulated scars to finalize the makeup and to reduce shine. For Model A, the following top detailing colors were used to create all three artificial scars (Mehron Inc., Chestnut Ridge,

New York): Mehron ProColoRing Bruise "Bloody Rose", Mehron ProColoRing Bruise "Midnight Sky", Mehron ProColoRing Bruise "Burnt Maroon". For Model B, Mehron ProColoRing Bruise "Midnight Sky" was used to create scar B1 (Figure 4.5) and scar B2 (Figure 4.6) whereas Mehron ProColoRing Bruise "Midnight Sky" and Mehron ProColoRing Bruise "Burnt Maroon" was used to create scar B3 (Figure 4.7)

4.2.2 Color measurements using a colorimeter

One of the widely used opto-electric tri-stimulus instruments commercially available for measuring skin colors in dermatology area is the Chroma-Meter (Konica Minolta, Osaka, Japan) [181-183]. Previous studies of the Chroma-Meter CR300 on normal skin have reported excellent intra-observer agreement (ICC ≥ 0.98) and good to excellent inter-observer agreement (ICC ≥ 0.85) for all of the three parameters (L*a*b*) [182].

A Chroma-Meter CR300 was used to record color information of specific regions of interest defined by a plastic surgeon (GPR). Before measurements were taken, the instrument was calibrated to a standard white plate provided by the manufacture. One measurement consisted of three consecutive flashes of illumination to obtain a mean value. Before any scar was created, color was measured between the breasts, on each breast mound, on the tan line (if applicable), in each nipple region, and of any natural scars. Only the skin color measured between the breasts were grouped and the mean value was taken as the control value for normal skin to which the color values of artificial scars and natural scars were compared.

For each artificial scar, color was measured on the scar region and multiple measurements were performed if the scar varied substantially in width. For accuracy and consistency of measurement, calibration was performed before a scar was created and

each time a new scar was created. A total of seven calibrations were used for the measurements. Past studies have shown that there are significant differences between the color parameters of hypertrophic scar regions (aged 6-12 months) due to burns, scalds, trauma and surgical incisions and normal skin regions, including upper arm, forearm, hand, chest, abdomen, back, and thigh [184]. The L* and b* values of normal skin collected using a spectro-colorimeter, Lab Scan XE (Hunter Associates Laboratory, Inc. Reston, VA), were significantly higher than those of hypertrophic scars, while measurements of a* of hypertrophic scars were significantly higher than those of normal skin. The authors did not report data on variability across racial groups in these trends of normal vs. scar measurements. To verify that our artificial scars also exhibit these properties, I analyzed the differences between normal skin regions and hypertrophic scars in CIE L*a*b*.

4.2.3 Color measurements using a camera

30 photographs, including before and after the artificial scars were applied, in sets of five standard views (AP, left oblique, left lateral, right oblique, right lateral), were taken. A Nikon 8400 was employed for all scar photographs because it is the model most commonly used for record keeping at The University of Texas M. D. Anderson Cancer Center. The main light source was overhead fluorescent lighting, and the camera was set to white balance, fluorescent lighting with ISO of 400. The aperture priority mode was set to the smallest value with spot metering for exposure level selection. Image quality was set to highest level possible to catch the details of the scars. The model stood in front of a sky blue background wall. Distance from subject to lens barrel was standardized to 150 cm.

For this study, scar regions of interest (ROIs) were delineated using the polygon selection tool of ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2007) by looking at the images at 400% zoom ratio on the computer screen with consistent resolution and size (1490×900). The RGB values were collected from each image for both the scar, and normal skin regions (between breasts). The reference values of "C" white used for the computations were provided by the Chroma-Meter CR300 instruction manual [185]. The CIE L*a*b* data of scar regions from photography were compared with those of colorimeter.

4.2.4 Statistical analysis

The agreement between color measurements obtained by colorimeter and digital photography was analyzed separately for each model. Three different statistical methods were used because each method has its technical limitations [186]. The methods used were: the Bland-Altman methods [174, 175], a hypothesis test for equivalence [172], and the ICC [173]. The Bland-Altman method provides a qualitative assessment while the hypothesis test for equivalence and ICC methods provide quantitative measures of agreement between measurements.

4.2.4.1 Bland-Altman method

Bland-Altman method is used for qualitatively measuring the degree of agreement, where the differences between photographic and colorimeter measurements are plotted against the average value of these measurements. If 95% of the differences are

within ±1.96 standard deviations of the mean of the differences, than the two methods are considered to be in good agreement.

4.2.4.2 Hypothesis test for equivalence

In testing for equivalence between two measurements, the null hypothesis is that the measurements by photography are not equivalent to those by colorimeter and the alternative hypothesis is that they are equivalent. This hypothesis is specifically intended for assessing equivalence and is distinct from that of the familiar paired *t*-test wherein a null hypothesis of equal value and an alternative hypothesis of unequal value are compared. The test statistic for assessing equivalence is

$$t = \frac{\sqrt{n}\left(\bar{x} \pm \delta\right)}{s}$$

where $\bar{x}$ and s denote the mean and standard deviation, respectively, of the differences between the photographic and colorimeter measurements. The value of δ is computed as a factor multiplied by the mean of the measurements by colorimeter. In this study, the factor was 0.20, 0.25, or 0.30. The variable δ accounts for the expected variability in the measurements made by two methods. A smaller value of δ implies stricter criteria for demonstrating that the measurements of the two methods are equivalent, in other words, photography can be used in place of colorimeter.

4.2.4.3 Intraclass correlation coefficient

To report the agreement between two measurements, a single-score two way random model was used since the set of images is a random subset of images from the

class of representative cases of post-operative images and the camera and colorimeter are also randomly selected from the population of cameras and colorimeters.

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where n denotes the number of scar measurements by photography and colorimeter and, k denotes the number of measurement methods, MS_R is the mean square error, MS_E is the residual mean square error, and MS_C is the mean square error between photographic and colorimeter measurements, respectively. The guideline for the interpretation of ICC used is that an ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility [187].

4.3 RESULTS

4.3.1 Comparison of artificial scars

Colorimeter measurements (CIE L*a*b*) of the artificial scars and normal skin were evaluated for characteristic differences (Table 4.5, Table 4.6) anticipated based on prior knowledge. Recall that the L*a*b* parameters indicate lightness, redness, and yellowness, respectively**.** The normal skin regions of model B had significantly higher a* and b* values than the normal skin regions of model A, while model A had a significantly higher L* value. This result is consistent with expectations since model B has a darker skin tone than Model A [188].

A similar pattern was seen when color measurements of the artificial scar regions were compared between the two models. Recall that the scars created for Model A resemble normal scars (Figure 4.2, Figure 4.3) and hypertrophic scars typically seen 3 to 6 months after surgery (Figure 4.4). The scars created for Model B resemble normal scars (Figure 4.5), hypertrophic scars (Figure 4.6) and keloid scar (Figure 4.7), typically seen 6 months to 1 year after surgery. The artificial scars of model B had substantially higher a* and b* values than the artificial scars of model A, while model A had a substantially higher L* value. This result is consistent with our expectations from clinical experience that scar colors vary across racial groups, but I am unaware of any systematic studies that have evaluated this trend. There were no notable differences between the L* and a* values of Model B's natural scar and those of the artificial scars created upon her skin. However, the artificial scars had lower b* values (less yellow) than her natural scar. I am unaware of any prior studies reporting how colorimeter measurements of scars change as scars mature. Thus, it is difficult to compare the measurements of the artificial scars, which were designed to emulate scars 6-12 months following surgery, to the subject's real scar, which was several years old.

Table 4.5 shows the colorimeter measurements of artificial scars created on Model A, and their comparison with those of her normal skin. Likewise, Table 4.6 shows the colorimeter measurements of artificial scars and normal skin of Model B. Artificial scars on model A had slightly lower L* values and substantially higher a* values than her normal skin. Similarly, artificial scars on model B also had slightly lower L* values and slightly higher a* values than her normal skin. In other words, as expected, the artificial scar regions look darker and redder than normal skin regions, as is the case for real scars. A previous study reported similar trends in the L* and a* measurements of real hypertrophic scars (6-12 months after skin injury) as compared with normal skin regions

[184]. In contrast, they reported that scars are less yellow (lower b* value) than normal skin [184]. However, there was no notable difference in the b* values of Model A's artificial scars and her normal skin or between the artificial scars and normal skin of Model B. Note that the authors [184] did not report data on variability across racial groups in these trends of normal vs. scar measurements.

The L*a*b* values of the artificial keloid scar (scar B3) were significantly lower than those of the artificial hypertrophic scars (scars B2). In other words, the artificial keloid scar was darker, less red, and less yellow than the artificial hypertrophic scars. This result is consistent with our expectations since hypertrophic scars should be lighter due to less melanin, redder due to the increase in vascularity, and less yellow as keloids are usually dark brown to black from clinical experience. However, I am unaware of any studies in which colorimeter measurements of keloid and hypertrophic scars were compared.

Table 4.5: Comparison between normal skin region and hypertrophic scar region of Model A in L*a*b* color space. L*a*b* parameters indicate lightness, redness, and yellowness, respectively. Group A scars showed slightly lower L* value and significantly higher a* value than normal skin. This result shows that artificial scar regions look darker and redder than normal skin regions, while there was no significant difference in b* value, which indicates no prominent yellowness between artificial scars and normal skin.

	Scar A1	Scar A2	Scar A3	Art. Scar Mean	Normal skin Mean
L*	74.75	53.72	70.33	66.27	67.00
a*	5.59	9.70	8.92	8.07	3.28
b*	6.00	17.78	13.72	12.50	11.12

Table 4.6: Comparison between normal skin region and hypertrophic scar region of Model A in L*a*b* color space. L*a*b* parameters indicate lightness, redness, and yellowness, respectively. There were no notable difference observed in L* and a* values between group B artificial scars and natural scar, except significantly and consistently lower b* in group B. There were no notable differences in all three L*a*b* values between natural scar and normal skin, because natural scar was 7 years aged according to Model B's history, and it was stabilized enough and didn't stand out against surrounding normal skin.

	Scar B1	Scar B2	Scar B3 (Keloid)	Art. Scar Mean	Natural scar Mean	Normal skin Mean
L*	41.02	64.07	34.43	46.51	46.74	49.28
a*	10.02	14.36	6.37	10.25	11.85	9.74
b*	10.55	10.84	6.63	9.34	17.89	19.68

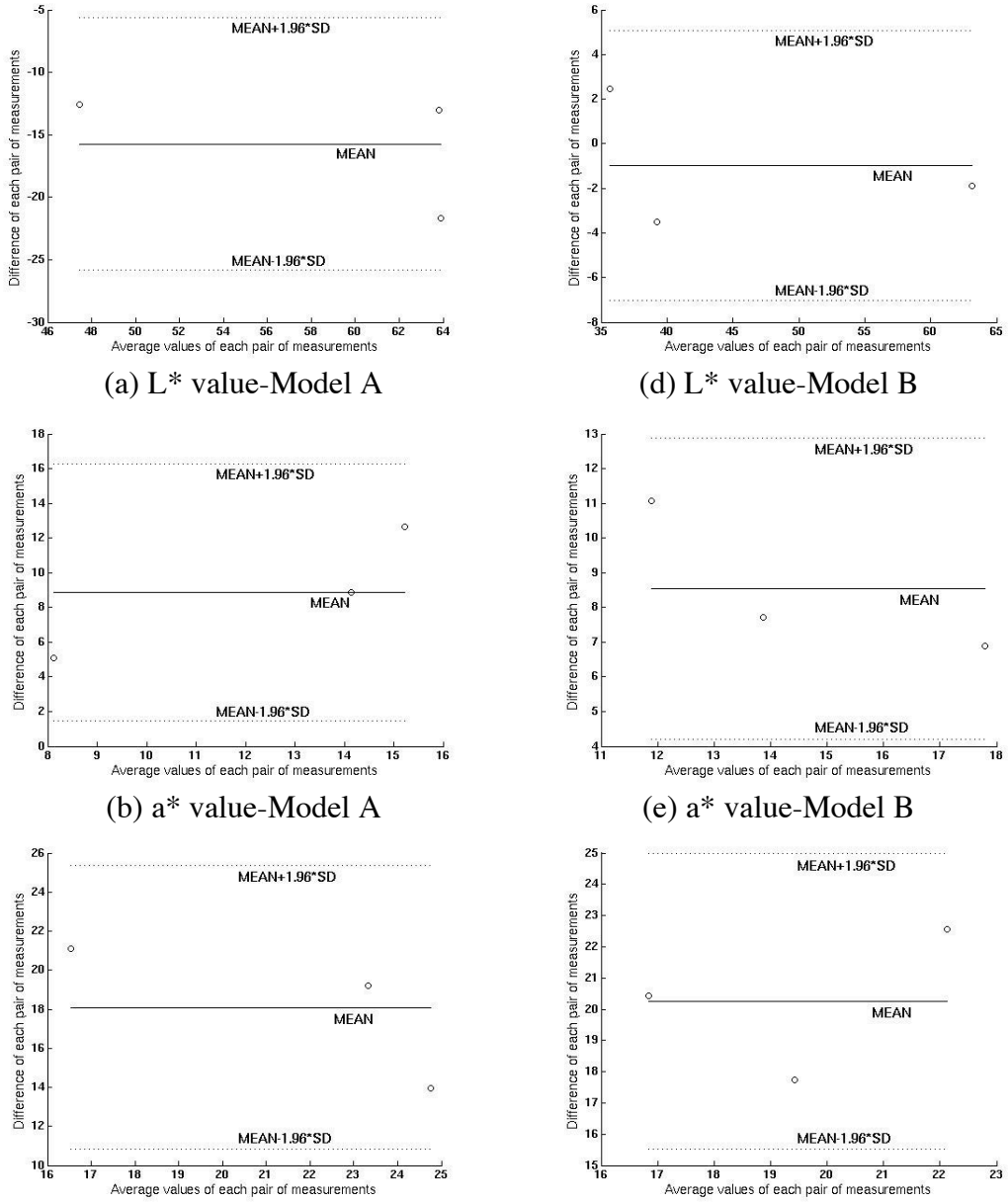
4.3.2 Inter-rater agreement

The agreement between photographic and colorimeter measurements was analyzed using the Bland-Altman method, a hypothesis test for equivalence, and ICC. Agreement was assessed separately for each model and was based on measurements of three artificial scars per model.

The degree of agreement between the measurements was also analyzed using the Bland-Altman method. Good agreement between the photographic and colorimeter measurements was obtained for all three parameters for both Models A and B as indicated by the fact that 95% of the differences were within the limit of agreement defined by Bland-Altman (Figure 4.8).

Using a hypothesis test for equivalence, agreement between the photographic and colorimeter measurements was demonstrated for both models at the 20% level for the a* ($p = 0.03$, $p = 0.03$) and b* ($p = 0.01$, p = 0.01) parameters (Table 4.7). Equivalence was also achieved at the 20% level for L* in Model B ($p = 0.04$), but was not achieved even at the 30% level for the L* value for Model A ($p = 0.08$).

Similarly, the degree of agreement between the measurements was assessed using the ICC (Table 4.8). ICC results were consistent with the hypothesis test for equivalence and Bland-Altman in that that there was “excellent” agreement between photographic and colorimeter measurements for the L* value of Model A (ICC = 0.99) and “fair to good agreement” for the L* value (ICC = 0.56) for Model B. In contrast to the other statistical assessments, ICC indicated unacceptable agreement in the a* and b* values (0.05–0.34)



(a) L* value-Model A (d) L* value-Model B

(b) a* value-Model A (e) a* value-Model B

(c) b* value-Model A (f) b* value-Model B

Figure 4.8: Bland-Altman analysis for the agreement of scar groups A (a, b, c) and B (d, e, f) between the measurements of three artificial scars (N = 3) for each of the three parameters that were measured by colorimeter and digital camera. The parameters measured were: (a) L value, (b) a* value, and (c) b* value. The results show that good agreement was obtained for all three parameters measured as indicated by the fact that 95% of the differences were within the limit of agreement defined by Bland-Altman.

Table 4.7: Results of the hypothesis test for equivalence of scar group B for three CIE L*a*b* parameter values of six artificial scars (N = 3) as measured by a camera and a colorimeter. Equivalence was achieved at 20% variability in all three color parameters except in group A, where equivalence was not achieved even at 30% in L* value. The null hypothesis was that the photographic and colorimeter color measurements are not equivalent. δ = factor*mean of colorimeter measurement. A *p*-value less than 0.05 indicates that the color measurements obtained by photography and colorimeter are equivalent.

Factor	L*		a*		b*	
	Group A	Group B	Group A	Group B	Group A	Group B
0.30	$p = 0.08$	$p = 0.02$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$
0.25	$p = 0.21$	$p = 0.02$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$
0.20	$p = 0.22$	$p = 0.04$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$

Table 4.8: Intra-class correlation (ICC) coefficient of group A and B for three CIE L*a*b* parameter values of six artificial scars as measured by a colorimeter and a camera. Mixed agreement was shown that group A (N = 3) showed "excellent" agreement for the L* value (ICC = 0.99), and "fair to good agreement" in group B for the L* value (ICC = 0.56), while Both group A and B showed unacceptable agreement in a* and b* values (0.05–0.34). The guideline for the interpretation of ICC used is that an ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

	L* value	a* value	b* value
Group A (N = 3)	ICC = 0.56	ICC = 0.34	ICC = 0.16
Group B (N = 3)	ICC = 0.99	ICC = 0.29	ICC = 0.05

4.4 DISCUSSION AND CONCLUSION

While colorimeters have been widely accepted as reliable tools for the assessment of skin color in Dermatology, they have not been routinely utilized in the Plastic and Reconstructive Surgery. However, digital photography is commonly used to document surgical outcomes. Thus, it would valuable if digital photography could also be used to quantitatively evaluate skin color. In order to assess the validity of color measurements obtained by clinical photography, I compared the measurements of a digital camera and a colorimeter. In order to conduct controlled experiments, artificial scars created by a makeup artist (TC) on professional models under the guidance of an experienced plastic surgeon (GPR) were used in this study. Comparisons between normal skin regions and artificial hypertrophic scars in CIE L*a*b* were consistent with expectations based on prior experience and previous studies [184]. Using multiple statistical methods, I demonstrated that color measurements obtained by digital photography were equivalent to those obtained using a colorimeter. Differences across the statistical methods may be due to the limited sample size (N=3 for each scar groups) I employed in this study. This study suggests that clinical photography can be employed for reliable, effective measurement methods of skin color for the analysis of surgical outcomes, but future, larger analyses employing real scars are needed.

While our results support the use of clinical photography for quantitative measurements of skin properties such as scarring, it is important to note that standardization of photographic conditions is critical. Substantial variability may be observed if the photographic conditions are not controlled. Important parameters to standardize include: lighting intensity, distance and direction from the light to the subject, camera lighting settings, background color, and the subject's pose relative to the camera.

While studio lighting is typically recommended for standardization of photography [121], in practice, this may not be feasible since it requires special equipment and training and that not be available to all clinicians. Fortunately, reliable images can be collected without studio conditions provided consistent settings are used. In this study, variability was controlled by avoiding the use of any built-in or external flashes and by simply setting the camera to white balance under fluorescent lamp, an option available on any commercial digital camera. This setting was also found to be advantageous in that it prevented a "washed-out" appearance of scars and surrounding skin colors.

Consistent with our previous study that employed manual identification of fiducial points on clinical photographs [189], I found that non-clinical observers could delineate regions of interest surrounding scars if a minimal amount of basic training was provided. However, a more automated method would be better for large outcome studies or for developing intra-operative tools since it took about 2 minutes to identify and delineate a region of interests.

In conclusion, I found that photography can be used in place of a colorimeter for measuring color properties of skin in a controlled setting. I also demonstrated that artificial scars created by a makeup artist under the guidance of a surgeon exhibit color properties consistent with real scars.

Chapter 5: ASSESSMENT OF SURGICAL SCARS FROM CLINICAL PHOTOGRAPHS

Contribution and publication: A preliminary version of the study in this chapter was presented at the American Medical Informatics Association Annual Symposium and a manuscript is being prepared for submission to a journal.
1. M. S. Kim, W. N. Rodney, J. Peng, G. P. Reece, and M. K. Markey, "Towards quantifying the aesthetic outcomes of breast cancer treatment: assessing surgical scars " in *American Medical Informatics Association Annual Symposium* Washington, D. C., 2005.

5.1 INTRODUCTION

The goal of reconstructive plastic surgery of the breast is to recreate a natural appearance that is satisfying to the patient. However, reconstruction surgery leaves scars that can be disfiguring, aesthetically unpleasant, and can strongly affect a patient's quality of life [151-153]. Scars can cause itching, tenderness, pain, sleep disturbances, anxiety, and depression. Other psychosocial problems include development of post-traumatic stress reactions, loss of self esteem, and stigmatization [151, 152]. The visibility of a scar is significantly influenced by surface area, vascularity, and pigmentation [156].

In the previous chapter, assessment of skin coloration by digital photography was validated through statistical analyses of measurements on artificial scars. In this chapter, real scars were evaluated using gray scale intensity values. The gray scale value of a pixel is the sum of all its RGB values. Grayscale processing eliminates the need to know the exact relationship between RGB and observed skin color and what proportion of the overall color is attributable to each component. Conversion of RGB to CIE L*a*b* color space requires the tri-stimulus values for the light source used. However, when taking routine clinical photographs, most plastic surgeons do not record the specific lighting

conditions that control the reference parameters. Thus, while we have demonstrated that color assessment by photography is valid, methods that employ grayscale intensity values only are also of practical value.

5.1.1 Surgical scar area assessment

The applicability of simple methods to measure the size of pathological skin lesions has been poorly studied to date. The inter-observer reliability and accuracy (validity) was established for planimetry (area of a 2-D figure) by tracing on a photograph and planimetry by tracing on a transparent sheet in a study by Zuijilen et al. [162]. Drawings of 25, 50, and $75\,cm^2$ were created on 3 locations with increasing curvature (back, thigh, and forearm) in 20 healthy volunteers. Three investigators evaluated the drawings by both planimetry techniques. Both techniques showed a good reliability using intraclass correlation coefficient (ICC ≥ 0.82) for $25\,cm^2$ areas. Planimetry by photography was more reliable than planimetry by transparent sheet for the 50 -and $75\,cm^2$ areas and was more accurate than planimetry by transparent sheet for all areas except for the area with the greatest curvature, the forearm. The study permits the conclusion that planimetry by photography is as suitable for surface area measurements as planimetry by transparent sheet.

In this study, we present results on using clinical photographs for the assessment of linear breast surgical scars. Digital clinical photographs were used to assess quantitative, objective measurements of breast surgical scars based on complete color intensity image analysis and area measurements. I demonstrate that the new objective measures utilizing clinical photographs are feasible, effective measurement methods using the Bland-Altman methods [191, 192]a hypothesis test for equivalence [190], and the ICC coefficient [173], and.

5.2 MATERIALS AND METHODS

5.2.1 Datasets

This retrospective analysis of clinical photography was approved by IRB protocols from The University of Texas at Austin and The University of Texas M. D. Anderson Cancer Center (UT MDACC). The patient population for this study was women aged 21 years or older who underwent Transverse Rectus Abdominis Myocutaneous (TRAM) breast reconstruction surgery from January 1, 1990 to June 1, 2003 at UT MDACC. The data pool consisted of digitized images from photographs taken by conventional 35 *mm* cameras and images from digital cameras. The conventional photographs were digitized with a Nikon LS 2000 or Nikon Super Coolscan 4000ED (1.06) slide digitizer. The digital cameras used were: Nikon 990 Coolpix or Cannon T90 35mm SLR with 50mm lens. All photographs were taken against a sky blue background, following guidelines on clinical photography in plastic surgery [121].

The data set for this study consisted of 40 post-operative anterior-posterior views of patients. The images were selected by an experienced plastic surgeon (GPR) and demonstrate a large variety of breast aesthetic characteristics, such as size, symmetry, ptosis, and projection. Out of the 40 anterior/posterior images used, 16 patients underwent conventional mastectomy followed by TRAM reconstruction and 24 patients had skin sparing mastectomy followed by TRAM reconstruction. Photographs of patients who undergo conventional mastectomy followed by TRAM (TRAM-CM) generally do not show the complete scar area because significant parts of scars are hidden below the breast mound or on the lateral side. Patients who undergo TRAM followed by a skin sparing mastectomy (TRAM-SSM) show no or very little flap or scarring in a standard photograph because the incision is made following the inframammary crease (Figure 5.1). All regions of interest (ROIs) were manually delineated using the polygon selection

tool in ImageJ (NIH) by looking at the images at 400% zoom ratio on the computer screen with consistent resolution and size (1490×900) across the study for obtaining the image information.

5.2.2 Measures of scar intensity (coloration) and area

Two kinds of measurements chosen for their significant impact on scar assessment were calculated for each image by two observers on the same image set on two occasions two weeks apart. The size of the scar was computed as a ratio of the area of the scar region to the total area of the affected breast (AR) based on the ImageJ delineation.

$$AR = \frac{A_{scar}}{A_{affected}} \tag{5.1}$$

$$NIG = \frac{|I_{scar} - I_{normal}|}{I_{normal}} \tag{5.2}$$

$$FIG = \frac{|I_{scar} - I_{flap}|}{I_{flap}} \tag{5.3}$$

The intensity gradient of the scar versus the surrounding normal skin was computed as the ratio of the absolute value of the difference in mean gray intensity of scar and surrounding normal regions to the average intensity of the surrounding normal region (NIG). This ratio provides a measure of the contrast difference between the two regions. For the TRAM-CM reconstructions, the intensity of the scar region was also

compared to that of the flap region (FIG) to investigate the contrast difference between the two regions.

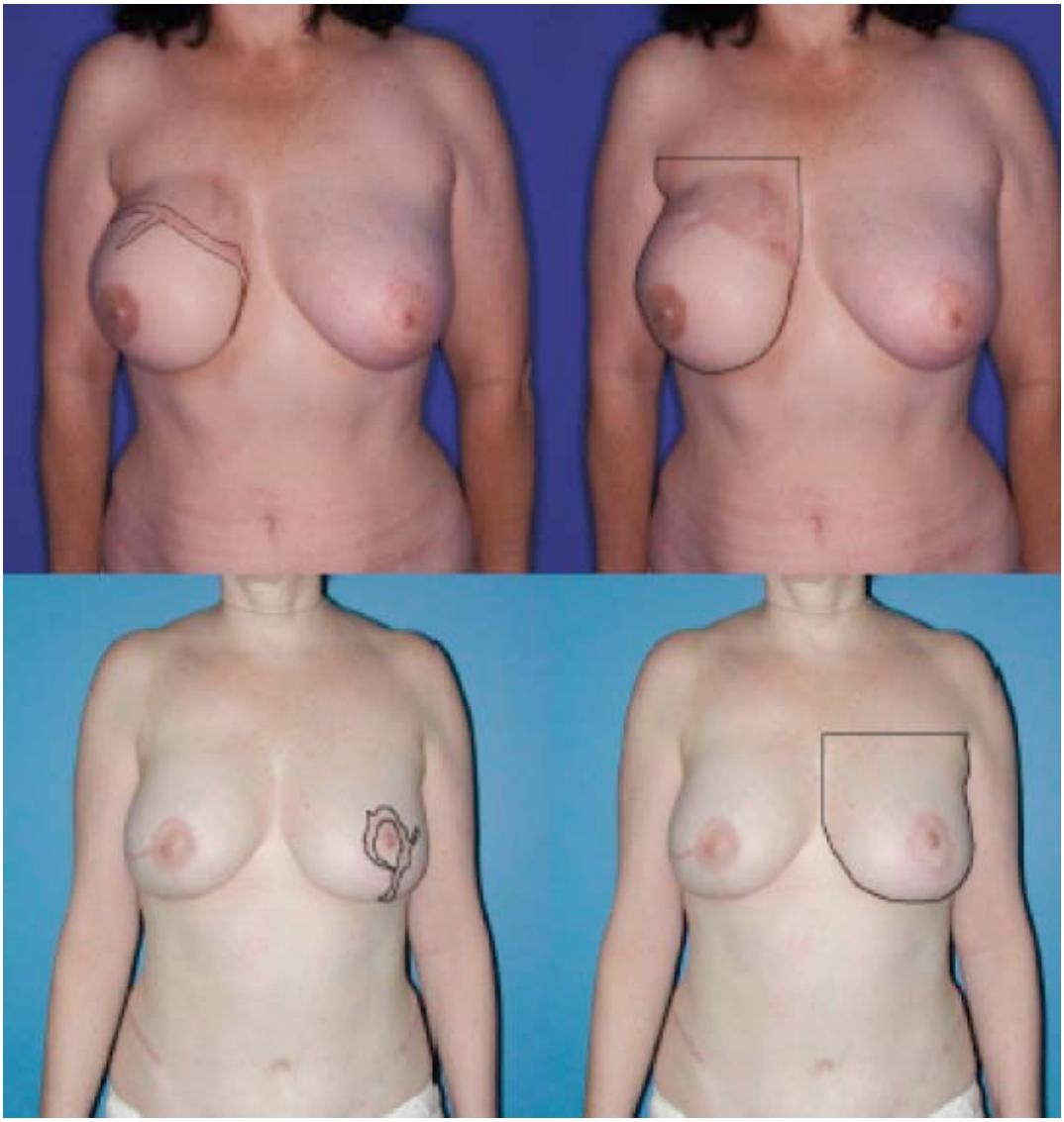


Figure 5.1: Sample images of regions of interest marked on clinical photographs. Top images show the areas of scar (left) and affected breast area (right) of patient who underwent conventional mastectomy followed by TRAM reconstruction. Bottom images show the areas of scar (left) and affected breast area (right) of patient who underwent skin-sparing mastectomy followed by TRAM reconstruction. For bottom images, only one of the patient's two affected breasts is marked in this figure for clear illustration.

5.2.3 Statistical analysis

The agreement between measurements obtained by two observers was analyzed separately for intensity and area. The methods used were: the Bland-Altman method [174, 175], a hypothesis test for equivalence [172], the ICC [173].. The Bland-Altman method provides a qualitative assessment while the hypothesis test for equivalence and ICC methods provide quantitative measures of agreement between measurements.

5.2.3.1 Bland-Altman method

The Bland-Altman method is used for qualitatively measuring the degree of agreement, where the differences in the measurements made by two measurements are plotted against the average value of these measurements. Good agreement is demonstrated by 95% of the differences being within ±1.96 standard deviations of the mean of the differences.

5.2.3.2 Hypothesis test for equivalence

In testing for equivalence between two measurements, the null hypothesis is that the measurements are not equivalent and the alternative hypothesis is that they are equivalent. This hypothesis is specifically intended for assessing equivalence and is distinct from that of the familiar paired *t*-test wherein a null hypothesis of equal value and an alternative hypothesis of unequal value are compared. The test statistic for assessing equivalence is

$$t = \frac{\sqrt{n}\left(\bar{x} \pm \delta\right)}{s}$$

where $\bar{x}$ and δ denote the mean and standard deviation, respectively, of the differences between the photographic and colorimeter measurements. The value δ is computed as a factor multiplied by the mean of the measurements by colorimeter. In this study, the factor was 0.20, 0.25, or 0.30. The variable δ accounts for the expected variability in the measurements made by two methods. A smaller δ value of implies stricter criteria for demonstrating that the measurements of the two methods are equivalent.

5.2.3.3 Intraclass correlation coefficient

To report the agreement within measurements by two observers, a single-score two way random model was used since the set of images is a random subset of images from the class of representative cases of post-operative images and observers are also randomly selected from the population of observers.

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where n denotes the number of object of measurement, k denotes the number of measurements, MS_R is the mean square error for between measurements of objects, MS_E is the residual mean square error, and MS_C is the mean square error between measurements.

An average-score two way random model ICC was used for analysis of inter-observer agreement. The average option was chosen because the averages of two measurements by each observer were used for within observer agreement.

$$ICC = \frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$$

where n denotes the number of images, MS_R is the mean square error between images, MS_E is the residual mean square error and MS_C is the mean square error between measurement methods. The guideline for interpretation of the ICC used is that an ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility [187].

The Bland-Altman method is used for qualitatively measuring the degree of agreement, where the differences in the measurements made by two measurements are plotted against the average value of these measurements. It is demonstrated that if 95% of the differences are within ±1.96 standard deviations of the mean of the differences, this denotes good agreements between two measurement methods.

5.2.4 Comparison of TRAM-CM and TRAM-SSM

In addition to these agreement assessments, two categories of data, TRAM-CM and TRAM-SSM, were analyzed for the area ratio and intensity measurements to investigate whether any characteristic differences between two groups are present.

5.3 RESULTS

The Bland-Altman method, a hypothesis test for equivalence, and ICC were used to assess the degree of intra- and inter-observer variability in measurements of scar

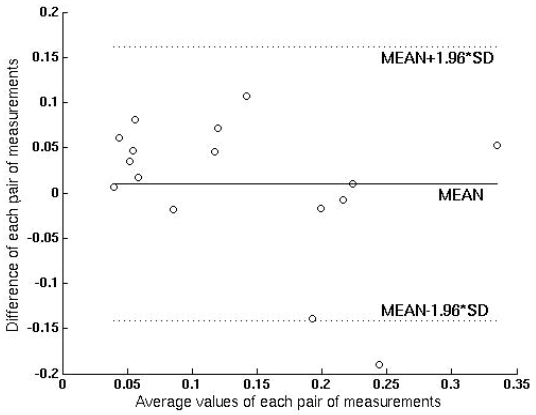
coloration and surface area due to variability in manual segmentation of regions of interest in digital photographs.

5.3.1 Intensity assessment, Intra-rater agreement

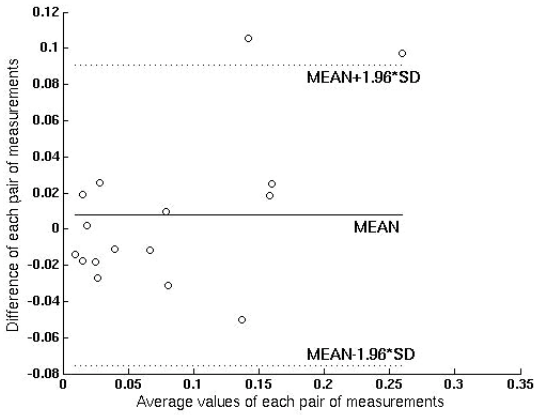
For both observers, the degree of agreement between the two sets of color measurements was analyzed using the Bland-Altman method (Figure 5.2, Figure 5.3). For both TRAM-CM and TRAM-SSM cases, good agreement was obtained for all the parameters measured; 95% of the differences were within the limit of agreement.

Using the hypothesis test for equivalence on the cases of TRAM-CM, equivalence was achieved between first and second color measurements (N = 16) of both observers at the 30% level for NIG (Observer 1: $p = 0.03$, Observer 2: $p = 0.03$) and at the 20% level for FIG (Observer 1: $p = 0.01$, Observer 2: $p = 0.01$), (Table 5.1). On the cases of TRAM-SSM, equivalence was achieved between the first and second measurements of NIG (N = 24) at the 20% level (Observer 1: $p = 0.00$, Observer 2: $p = 0.00$) (Table 5.2).

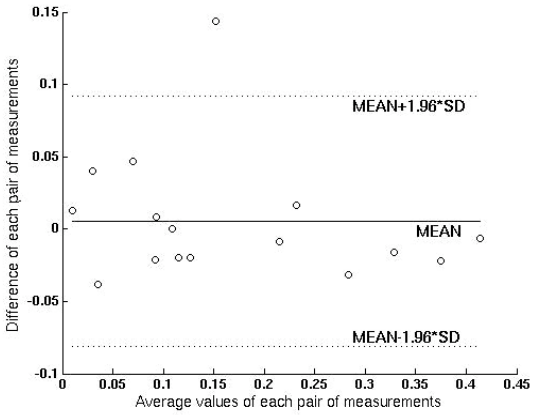
The intra-observer agreement as assessed by ICC between the two sets of measurements made by Observer 1 on the cases of TRAM-CM was very good (Table 5.3). For Observer 1, the intra-observer agreement was “excellent” for FIG (ICC = 0.94) and “fair to good” for NIG (ICC = 0.68). Similarly, for Observer 2, the intra-observer agreement was “excellent” for both NIG (ICC = 0.84) and FIG (ICC = 0.95). Likewise, the intra-observer agreement on the cases of TRAM-SSM was very good (Table 5.4). The intra-observer agreement for NIG on the TRAM-SSM cases was “fair to good” for Observer 1 (ICC = 0.61) and “excellent” for Observer 2.



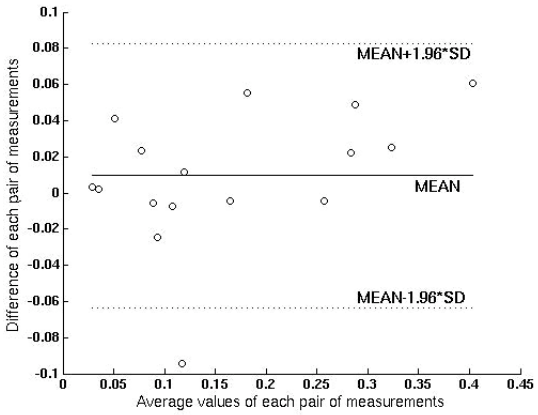
(a) NIG – Observer 1



(b) NIG - Observer 2

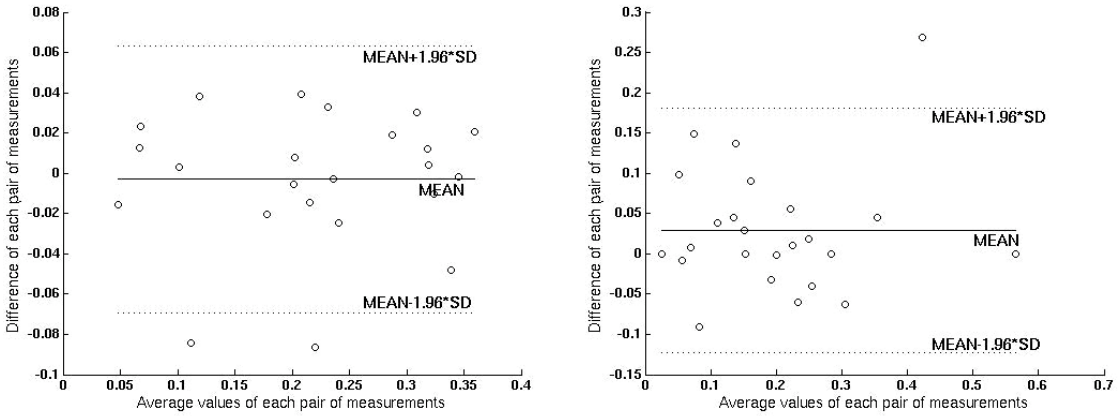


(c) FIG – Observer 1



(d) FIG – Observer 2

Figure 5.2: Bland-Altman analysis for the intra-observer agreement for the intensity measurement ratios that were measured by twice by the Observer 1 and Observer 2 on the cases of TRAM after conventional mastectomy (TRAM-CM). The parameters measured were: (a) NIG-Observer 1, (b) NIG-Observer 2, (c) FIG-Observer 1, and (d) FIG-Observer 2. Good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.



(a) NIG – Observer 1 (b) NIG – Observer 2

Figure 5.3: Bland-Altman analysis for the intra-observer agreement for the intensity measurement ratios that were measured by twice by Observer 1 and Observer 2 on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). The parameters measured were: (a) NIG – Observer 1, and (b) NIG - Observer 2. Good agreement was obtained for the parameter measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Table 5.1: Results of the hypothesis test for equivalence between the first and second set of measurements made by Observer 1 and Observer 2 on the cases of TRAM after conventional mastectomy (TRAM-CM). The null hypothesis was that the two sets of measurements made by Observer1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the two sets of measurements. δ = Variability factor*mean of first set of measurement.

Number of cases	Variability Factor	Observer 1		Observer 2	
		NIG	FIG	NIG	FIG
16	0.30	$p = 0.03$	$p = 0.00$	$p = 0.03$	$p = 0.00$
16	0.25	$p = 0.06$	$p = 0.00$	$p = 0.06$	$p = 0.00$
16	0.20	$p = 0.11$	$p = 0.01$	$p = 0.11$	$p = 0.01$

Table 5.2: Results of the hypothesis test for equivalence between the first and second set of measurements made by Observer 1 and Observer 2 on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). The null hypothesis was that the two sets of measurements made by Observer1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the two sets of measurements. δ = Variability factor*mean of first set of measurement.

Number of cases	Variability Factor	Observer 1	Observer 2
		NIG	NIG
16	0.30	$p = 0.00$	$p = 0.00$
16	0.25	$p = 0.00$	$p = 0.00$
16	0.20	$p = 0.00$	$p = 0.00$

Table 5.3. Intraclass correlation (ICC) coefficient for NIG and FIG ratios between the first and second set of measurements made by each observer on the cases of TRAM with conventional mastectomy (TRAM-CM). The intra-observer agreement was "excellent" for Observer 1's FIG, Observer 2's NIG (ICC = 0.84), and FIG (ICC = 0.95). "Fair to good" agreement was achieved for Observer 1's NIG (ICC = 0.68) values. ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

Number of cases (N =16)	NIG	FIG
Observer 1	ICC = 0.68	ICC = 0.94
Observer 2	ICC = 0.84	ICC = 0.95

Table 5.4: Intraclass correlation (ICC) coefficient for AR and NIG ratios between the measurements made by each observer on the cases of TRAM with a skin sparing mastectomy (TRAM-SSM). The intra-observer agreement was "excellent" for Observer 2's NIG (ICC = 0.83), while the reproducibility for Observer 1's NIG (ICC = 0.32) was "poor". ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

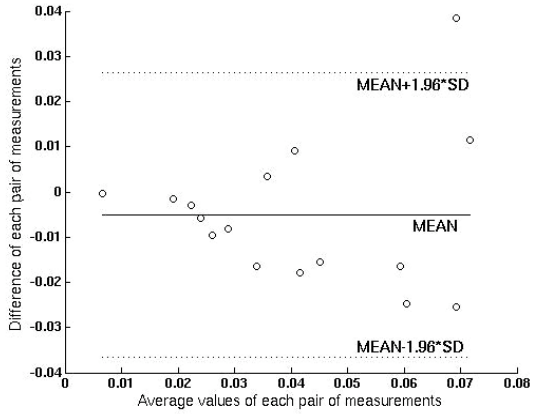
Number of cases (N = 24)	NIG
Observer 1	ICC = 0.61
Observer 2	ICC = 0.83

5.3.2 Area Assessment, Intra-rater agreement

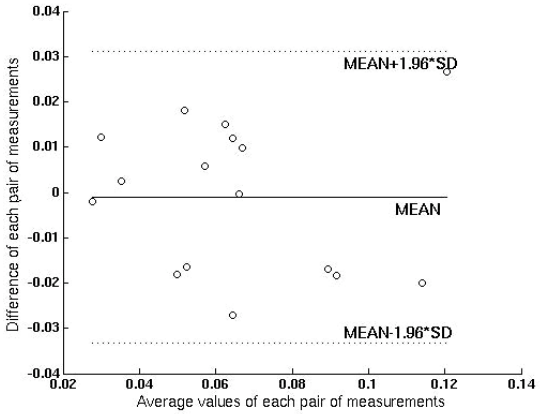
For both observers, the degree of agreement between the two sets of area measurements was analyzed using the Bland-Altman method (Figure 5.4, Figure 5.5). For both TRAM-CM and TRAM-SSM cases, good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Using the hypothesis test for equivalence on the TRAM-CM cases, equivalence was achieved between first and second area measurements (N = 16) at the 30% level for Observer 1 (p = 0.03) and at the 20% level for Observer 2 (p = 0.00) (Table 5.5). For an equivalence level of 20%, equivalence was achieved between the first and second measurements (N = 24) of both observers for the measurement ratios on the cases of TRAM-SSM: Observer 1's AR ($p = 0.01$), Observer 2's AR ($p = 0.00$), (Table 5.6).

The intra-observer agreement between the two sets of area measurements was evaluated by ICC. For the area measurements made by Observer 1 on the cases of TRAM-CM, the agreement was "fair to good" (ICC = 0.72). The ICC was "excellent" for Observer 2 (ICC = 0.83), (Table 5.7). The intra-observer agreement between the two sets of measurements made by each observer on the cases of TRAM-SSM using the ICC coefficient showed mixed variability (Table 5.8). The intra-observer agreement was "good" for Observer 2's AR (ICC = 0.63), while the reproducibility for Observer 1's AR was "poor" (ICC = 0.18).

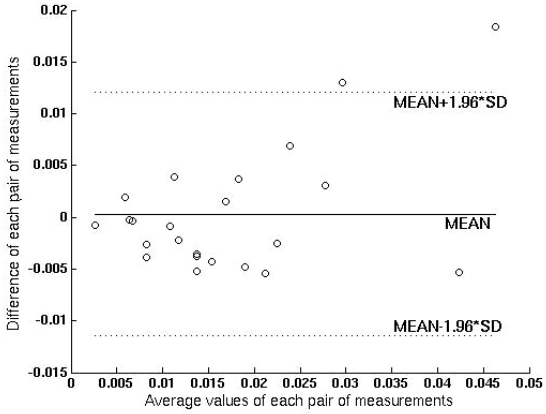


(a) AR - Observer 1

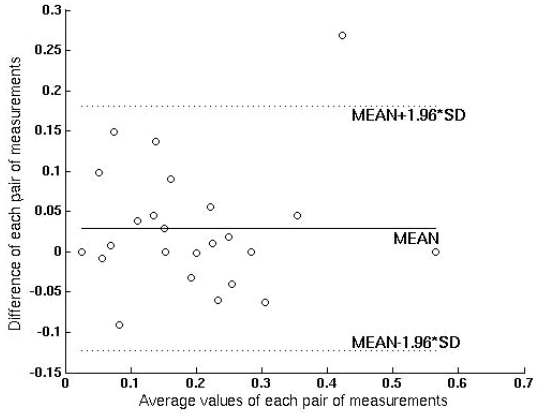


(b) AR - Observer 2

Figure 5.4: Bland-Altman analysis for the intra-observer agreement for area ratios that were measured by twice by each observer on the cases of TRAM after skin sparing mastectomy (TRAM-CM). The parameters measured were: (a) AR - Observer 1, (b) AR - Observer 2. Good agreement was obtained for the parameter measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.



(a) AR - Observer 1



(c) AR - Observer 2

Figure 5.5: Bland-Altman analysis for the intra-observer agreement for each of the two scar measurement ratios that were measured by twice by each observer on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). The parameters measured were: (a) AR-Observer 1, (b) AR-Observer 2, Good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Table 5.5: Results of the hypothesis test for equivalence between the first and second set of measurements made by Observer 1 and Observer 2 on the cases of TRAM after conventional mastectomy (TRAM-CM). The null hypothesis was that the two sets of measurements made by Observer1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the two sets of measurements. δ = Variability factor*mean of first set of measurement.

Number of cases	Variability Factor	Observer 1 AR	Observer 2 AR
16	0.30	$p = 0.03$	$p = 0.00$
16	0.25	$p = 0.07$	$p = 0.00$
16	0.20	$p = 0.13$	$p = 0.00$

Table 5.6: Results of the hypothesis test for equivalence between the first and second set of measurements made by Observer 1 and Observer 2 on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). The null hypothesis was that the two sets of measurements made by Observer1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the two sets of measurements. δ = Variability factor*mean of first set of measurement.

Number of cases	Variability Factor	Observer 1 AR	Observer 2 AR
16	0.30	$p = 0.00$	$p = 0.00$
16	0.25	$p = 0.00$	$p = 0.00$
16	0.20	$p = 0.01$	$p = 0.00$

Table 5.7. Intraclass correlation (ICC) coefficient for AR, NIG, and FIG ratios between the first and second set of measurements made by each observer on the cases of TRAM with conventional mastectomy (TRAM-CM). The intraobserver agreement was "excellent" for Observer 2's AR (ICC = 0.83). "Fair to good" agreement was achieved for Observer 1's AR (ICC = 0.72) values. ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

Number of cases (N =16)	AR
Observer 1	ICC = 0.72
Observer 2	ICC = 0.83

Table 5.8: Intraclass correlation (ICC) coefficient for AR and NIG ratios between the measurements made by Observer 1 and Observer 2 on the cases of TRAM with a skin sparing mastectomy (TRAM-SSM). The intra-observer agreement was "good" for Observer 2's AR (ICC = 0.63), while Observer unacceptable reproducibility was achieved for Observer 1's AR (ICC = 0.18). ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

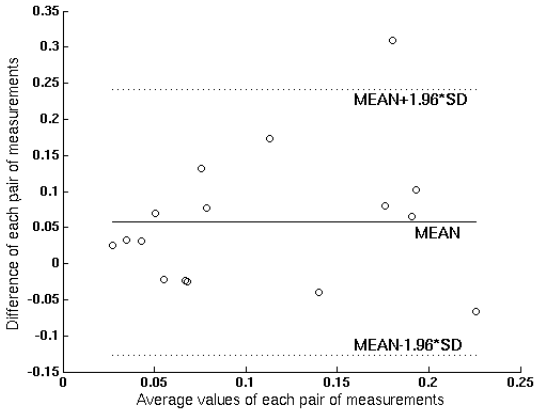
Number of cases (N = 24)	AR
Observer 1	ICC = 0.87
Observer 2	ICC = 0.63

5.3.3 Intensity assessment, Inter-rater agreement

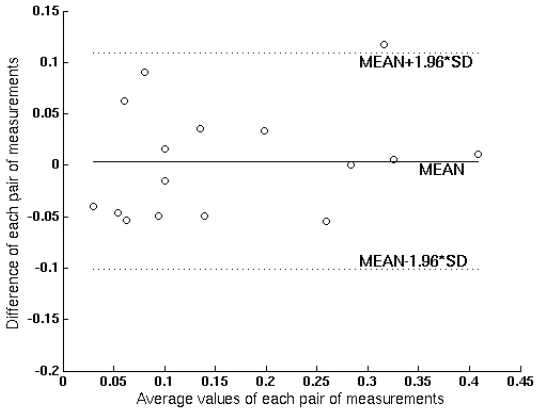
The inter-observer agreement in the color assessments made based on the manual segmentations by the two observers was analyzed using the Bland-Altman method (Figure 5.6). For both TRAM-CM and TRAM-SSM cases, good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Using the hypothesis test for equivalence to assess the inter-observer variability in NIG for TRAM-CM cases (N = 16), the results were only borderline for statistical significance at the 30% level ($p = 0.05$). In comparison, inter-observer equivalence was achieved for FIG on the TRAM-CM cases at the 20% level ($p = 0.01$). For the TRAM – SSM (N = 24) cases, inter-observer equivalence in NIG was achieved at the 30% level ($p = 0.03$) (Table 5.9).

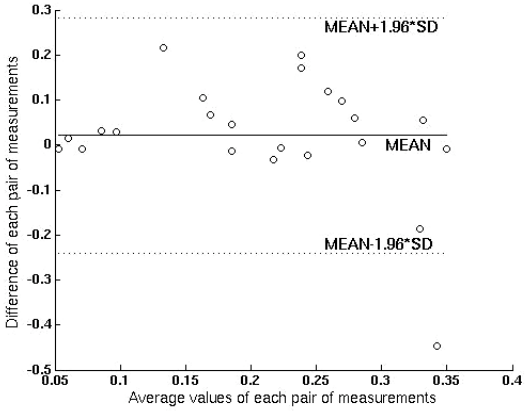
The inter-observer agreement between the color measurements made by Observer 1 and Observer 2 on the cases of TRAM-CM was also assessed using ICC. The inter-observer agreement on TRAM-CM cases was "fair to good" for NIG (ICC = 0.50) and "excellent" for FIG (ICC = 0.95). Similarly, for TRAM-SSM cases, the inter-observer agreement was "fair to good" for NIG (ICC = 0.62) (Table 5.10).



(a) TRAM-CM, NIG



(b) TRAM-CM, FIG



(c) TRAM-SSM, NIG

Figure 5.6: Bland-Altman analysis for the inter-observer agreement between Observer 1 and Observer 2 for each of the three scar measurement ratios that were measured on the cases of TRAM after conventional mastectomy (TRAM-CM) and on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). The parameters measured were: (a) TRAM-CM, NIG, (b) TRAM-CM, FIG, and (c) TRAM-SSM, NIG. Good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Table 5.9: Results of the hypothesis test for equivalence between the observers on the cases of TRAM with conventional mastectomy (TRAM-CM) and on the cases of TRAM after a skin-sparing mastectomy (TRAM-SSM). The null hypothesis was that the measurements made by Observer1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the measurements by two observers. δ = Variability factor*mean of first set of measurement.

	TRAM-CM (N=16)		TRAM-SSM (N=24)
Variability factor	NIG	FIG	NIG
0.30	$p = 0.05$	$p = 0.00$	$p = 0.03$
0.25	$p = 0.08$	$p = 0.00$	$p = 0.06$
0.20	$p = 0.12$	$p = 0.01$	$p = 0.11$

Table 5.10: Intraclass Correlation Coefficient (ICC) for NIG and FIG measurement ratios between Observer 1 and Observer 2 on the cases of TRAM with conventional mastectomy (TRAM-CM) and on the cases of TRAM with skin sparing mastectomy (TRAM-SSM). ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

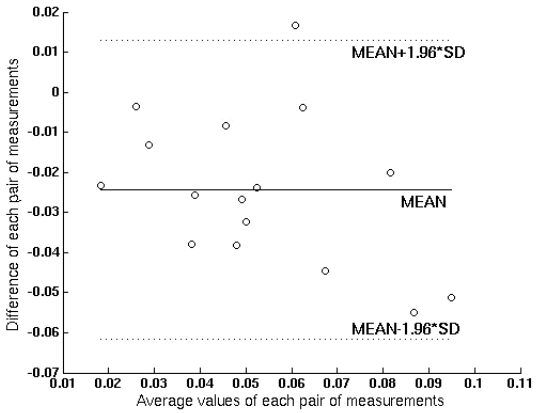
TRAM-CM (N=16)		TRAM-SSM (N=24)
NIG	FIG	NIG
ICC = 0.50	ICC = 0.95	ICC = 0.62

5.3.4 Area assessment, Inter-rater agreement

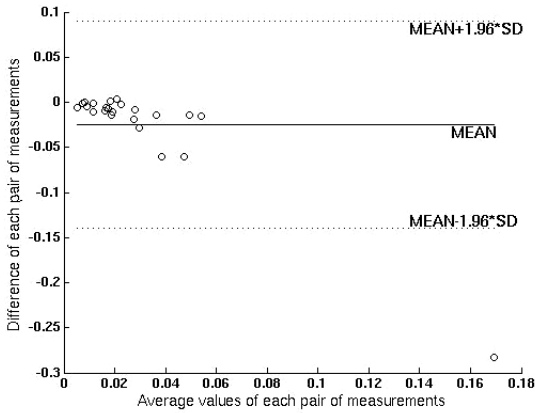
The inter-observer agreement in the area measurements made based on the manual segmentations of the two observers was analyzed using the Bland-Altman method (Figure 5.7). For both TRAM-CM and TRAM-SSM cases, good agreement was obtained for all the parameters measured because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

The inter-observer agreement in the area measurements was also evaluated using the hypothesis test for equivalence (Table 5.11). For the TRAM-CM cases (N = 16), equivalence between the areas measurements based on the ROIs identified by the two observers was achieved at the 20% level ($p = 0.00$). Likewise, for the TRAM-SSM cases (N = 24), equivalence between the areas measurements based on the ROIs identified by the two observers was achieved at the 20% level ($p = 0.04$).

As evaluated by ICC, the inter-observer agreement between the area measurements made by Observer 1 and Observer 2 on the cases of TRAM-CM was "excellent" (ICC = 0.81) and FIG (ICC = 0.95) (Table 5.12). "Fair to good" agreement in the area measurements was achieved for the TRAM-SSM cases (ICC = 0.68).



(a) TRAM-CM, AR



(b) TRAM-SSM, AR

Figure 5.7: Bland-Altman analysis for the inter-observer agreement in AR between Observer 1 and Observer 2 on the cases of TRAM after skin sparing mastectomy (TRAM-SSM) and on the cases of TRAM after skin sparing mastectomy (TRAM-SSM). Good agreement was obtained because 95% of the differences were within the limit of agreement following Bland-Altman's suggestion.

Table 5.11: Results of the hypothesis test for equivalence between the observers on the cases of TRAM after conventional mastectomy (TRAM-CM) and on the cases of TRAM after a skin-sparing mastectomy (TRAM-SSM). The null hypothesis was that the measurements made by Observer 1 and Observer 2 were not equivalent. A *p*-value less than 0.05 indicate that equivalence can be achieved between the measurements by two observers. δ = Variability factor*mean of first set of measurement.

	TRAM-CM (N=16)	TRAM-SSM (N=24)
Variability factor	AR	AR
0.30	$p = 0.00$	$p = 0.03$
0.25	$p = 0.00$	$p = 0.03$
0.20	$p = 0.00$	$p = 0.04$

Table 5.12. Intraclass correlation (ICC) coefficient for AR measurement ratios between Observer 1 and Observer 2 on the cases of TRAM with conventional mastectomy (TRAM-CM) and on the cases of TRAM with skin sparing mastectomy (TRAM-SSM).

TRAM-CM (N=16)	TRAM-SSM (N=24)
AR	AR
ICC = 0.81	ICC = 0.68

5.3.5 Comparison of TRAM-CM and TRAM-SSM

In addition to the agreement assessments, two categories of data, TRAM-CM and TRAM-SSM, were analyzed for the area ratio and intensity measurements to investigate whether there present any characteristic differences between two groups (Table 5.13). One reason for these differences is that the identification of an ROI of "normal" skin seemed to be more variable than for identifying "flap" skin. Two-sample t-test was performed to analyze the difference between the ratio values of these two groups. The results of AR measurements showed that the proportion of the breast area affected by scarring is statistically significantly larger (p = 0.01) for TRAM-CM (overall mean = 0.05) than for TRAM-SSM (overall mean: 0.03). This result is consistent with our expectation that larger skin flap introduce larger scar regions. The results of intensity measurements showed that intensity gradient between the scars and surrounding normal scars were statistically significantly greater (p = 0.00) in TRAM-SSM (overall NIG = 0.21) than that of TRAM-CM (overall NIG = 0.11). When the two intensity measurements within TRAM-CM were compared, the intensity gradient between scars and flap (overall FIG = 0.17) were greater than that between scars and surrounding normal skin area (overall NIG = 0.11, but the result was not statistically significant (p = 0.78). The possible explanation for this result is that significant amount of scars on the cases of TRAM-CM, most of the breast skin is replaced with transported skin from abdominal area which were not affected by breast cancer treatments such as radiation therapy which may make skin appear tan or red, while most of the skin is preserved on the cases of TRAM-SSM.

Table 5.13: Results of the AR values by the observers on the cases of TRAM after a conventional mastectomy (TRAM-CM, N = 16) and on the cases of TRAM after a skin-sparing mastectomy (TRAM-SSM, N = 24).

Observer	TRAM-CM (N=16)			TRAM-SSM (N=24)	
	AR	NIG	FIG	AR	NIG
Observer 1	0.04	0.14	0.17	0.02	0.22
Observer 2	0.07	0.08	0.16	0.04	0.20
Overall	0.05	0.11	0.17	0.03	0.21

5.4 DISCUSSION AND CONCLUSION

Breast surgery leaves scars that can be disfiguring, aesthetically unpleasant, and can strongly affect a patient's quality of life. Current assessments of surgical scars require employment of mechanical, electronic, or chemical measurements that are expensive and time consuming or employment of qualitative observer ratings based on direct viewing of patients and which have low intra- and inter- observer agreement. Quantitative as well as effective measurement methods are required to investigate the relationships between patient and surgical variables and aesthetic outcomes.

In this chapter, I presented results of using clinical photographs for the assessment of color and area information of linear breast surgical scars. Digital photography has already been widely utilized in the plastic and reconstructive surgery field for record keeping purposes. I investigated intra-and inter-reliability of the measurements in TRAM-CM and TRAM-SSM groups using the Bland-Altman method [191, 192], a hypothesis test for equivalence [190], the ICC coefficient [173]. The results showed no statistical difference between and within two non-clinical observers in the two measures using photographs of real scars.

Based on the fair to excellent reliability for most of the measurement methods achieved in the assessment, two categories of data, TRAM-CM and TRAM-SSM, were

also analyzed s to see if there are any characteristic differences in the area ratio and intensity measurements between the two groups. The AR measurement suggests that breast area affected by scarring is larger for TRAM-CM than for TRAM-SSM, which is consistent with expectations due to the skin-sparing procedure using a smaller flap in reconstruction and involving a smaller perimeter. The results of intensity measurements of the TRAM-CM show that the intensity gradient between scar and flap is greater than between scar and normal skin. This result implies that the scar appears more visible when compared to the flap skin than when compared to the normal skin in the area. However, the TRAM-SSM showed larger intensity gradients between scar and normal skin than those found for the scar and flap of TRAM-CM.

Two important points for future studies of surgical scar assessment by digital photography should be noted. First, standardized photography is critical for reliable photographic assessment of scars. Even though the data used in this study were collected following the current photographic standards [121], the authors still had difficulty in delineating scar regions on some patient images because of varying lighting intensity, distance and direction from the light to the subject, camera lighting settings, background color, and a subject's pose in relation to a camera. Second, all pertinent information on the photographic conditions, including the tri-stimulus values of the light source, should be recorded. Likewise, thorough clinical information accompanying the clinical photographs will help clinicians understand the measurement results.

This study presented encouraging findings and suggests that objective quantitative measures for assessing surgical scars can be computed. Further study employing large patient cases will be necessary to confirm the results of this study.

Chapter 6: DEVELOPMENT OF OBSERVER RATING SCALE FOR THE ASSESSMENT OF BREAST AESTHETICS

Contribution and publication: The study in this chapter is ongoing under the guidance of an experienced behavioral scientist Dr. Basen-Engquist (Behavioral Science, UT MDACC).

6.1 INTRODUCTION

While a human observer's rating assessment of the aesthetic outcome is not as reliable as an objective measure would be, it is the method most commonly used at this time (Chapter 2). However, although prior studies have used an observer rating scale for breast aesthetics, there is not a standard, accepted scale. Moreover, there are problems with the existing scales. First, most of the scales are extremely vague. Second, most focus only on one or two aesthetic properties (*e.g.*, symmetry) and so provide a very limited measure of aesthetics. Third, most of the scales were devised with breast conservation therapy in mind and can't be directly used to assess the outcomes of other surgeries, such as reconstruction procedures.

We are developing an observer rating scale following social science guidelines for scale construction [193-195], under the guidance of an experienced behavioral scientist Dr. Basen-Engquist (Behavioral Science, UT MDACC). The purpose is to enable plastic surgeons to subjectively, but reliably, describe breast anatomy. The wide-spread adoption of the lexicon would help standardize the terminology used to report the aesthetic outcomes of reconstructive surgery and improve communication between reconstructive surgeons and radiation oncologists.

6.2 FRAMEWORK FOR RATING SYSTEM DEVELOPMENT

While many previous studies have employed a single-item overall assessment of aesthetics, a multi-item scale is typically recommended for complex phenomena. We began by generating a large pool of potential items through "brainstorming" sessions. At the same time, we discussed possible question formats and the number of response categories. After generating a pool of items, we rated them for relevance and clarity so as to choose a reasonable number for use. We collected audio recordings of our scale construction meetings so that we can answer questions that may later arise about the scale construction process.

Two important properties of a scale are validity and reliability [193-195]. The content validity of our rating system is established by the involvement of reconstructive surgeons throughout the scale construction process. We are quantifying the reliability of our rating system using the Intraclass Correlation Coefficient (ICC).

Important design criteria for producing reliable subjective scales are clarity and explicitness [193]. The lexicon of breast aesthetic terminology and assessment format is meant to standardize the language used to report the observer assessment of aesthetic outcomes in reconstructive surgery and improve communication between reconstructive surgeons and radiation oncologists. Thus, the descriptions of the levels of the scale items will include text, idealized line drawings, and actual photographic examples. A preliminary version of the lexicon is presented in section 6.3.

6.2.1 Data collection

Three experienced plastic surgeons at UT MDACC used the preliminary lexicon to rate 32 patient cases using a web-based interface. This web-based interface was also used to display the clinical images and collect observer ratings of aesthetics during the

eye-tracker experiments (Figure 6.1). Five images of each patient are displayed: an anterior-posterior and lateral and oblique views of both the right and left breasts. The oblique views are taken at an angle of about 45 degrees from the true lateral. The vertical extent of the views is from just below the chin down to the top of the pubic bone. The oblique views are taken at an angle of about 45 degrees from the true lateral. All images were displayed on a LCD monitor using the Internet Explorer (© Microsoft Corporation) browser.

6.2.2 Evaluation of preliminary lexicon

In addition to validity and reliability, it is important to confirm that different observers can consistently use our rating system. Repeatability is being assessed by determining the agreement in using the rating system between different observers and over time. As an initial step toward to this evaluation process, six fellows at UT MDACC who weren't involved in designing the rating system rated 13 selected cases using the same web-based interface the plastic surgeons used. ICC was used to measure intra- and inter-rater agreement. We standardized the reading conditions as much as possible. For example, all of the observers viewed the cases using the same computer facility in the department.

In addition to assess repeatability of the scale, we have been evaluating the lexicon to generate explicit terminology. As a preliminary evaluation process, we found that items about breast shape are important. Two plastic surgeons rated the top-six most influential rating items selected from a preliminary breast aesthetic rating instrument for eye-tracking study (Refer Ch 7). The goal of this study was to understand how plastic surgeons assess breast aesthetics by using eye-tracking technology to record their gaze path while they rate breast anatomy on clinical photographs. The rating items used were:

(1) initial impression of overall appearance of the breasts, from 0 indicating very poor to 10 indicating excellent, (2) symmetry of size of breast mounds (3) symmetry of shape of breast mounds, (4) aesthetic shape (5) natural shape and (6) final impression of overall appearance of the breasts.

More work should be done to develop explicit observer rating scale to rate full range of breast aesthetics based on patients' opinions as well as the knowledge of experienced reconstructive surgeons. Both the observer rating scale and quantitative methods should be assessed for the intra- and inter-observer variability against each other using rating data collected by more reconstructive surgeons and patients not involved in designing the scale.

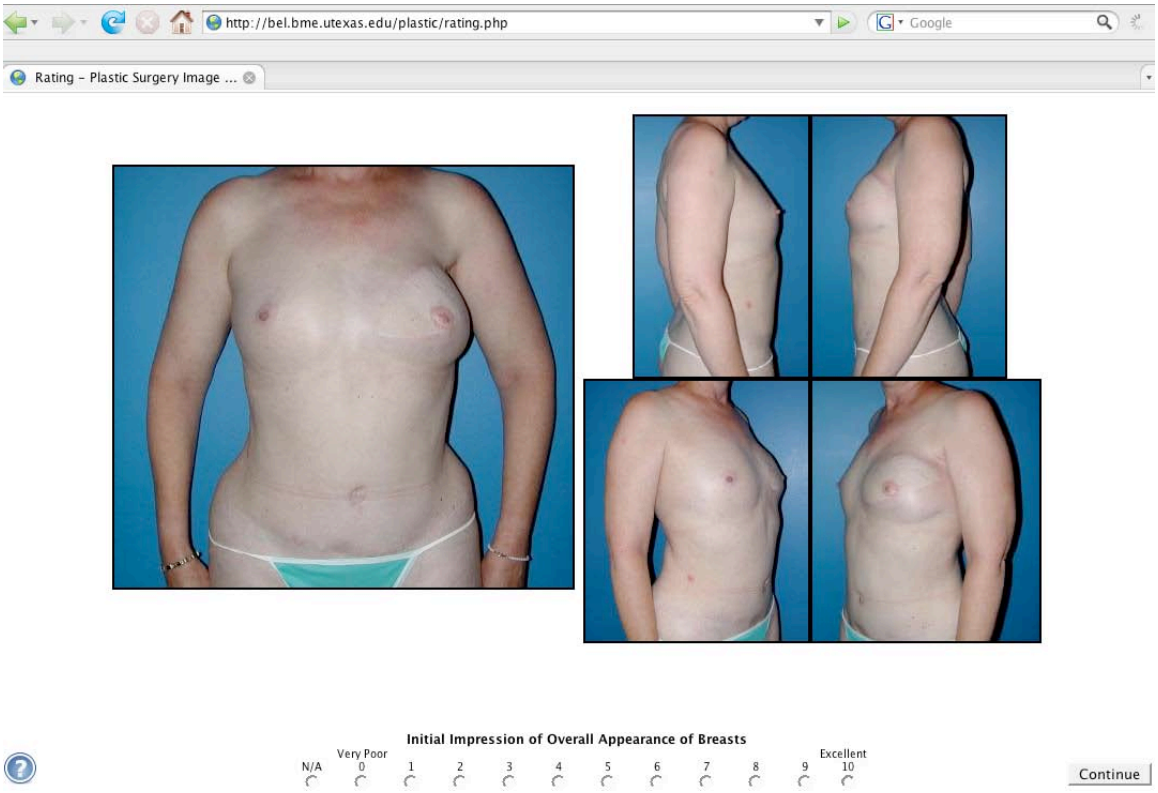


Figure 6.1: This image illustrates an example of a patient case displayed while an observer examined a case to rate the initial impression of the overall appearance of the breasts. The observer rated this case on the 11-point scale, where 0 indicates very poor and 10 indicates excellent. Five images are displayed from each patient: an anterior-posterior (AP) and lateral and oblique views of both the right and left breasts. A specific item is displayed at the same time the images are displayed. Clicking on question mark at the bottom left corner gives help files reproduced in section 6.3.

6.3 PRELIMINARY OBSERVER RATING SCALE

The observer rating scale consists of 11 symmetry ratings items, 14 individual breast ratings items, and a global rating on overall appearance before and after the entire rating items. To rate the symmetry rating items, observer should select a rating from a point scale or Y/N. If the observer feel this item cannot be properly rated, he/ she can select “N/A”. Similarly, observer should select a rating from a point scale or Y/N, but should rate both breasts separately. Each rating scale is explained with the lexicon prepared to help standardize the terminology used to report the subjective assessment of aesthetic outcomes in reconstructive surgery and improve communication between reconstructive surgeons and radiation oncologists.

6.3.1 Impression of overall appearance of breasts

	Very Poor								Excellent
N/A	0	1	2	3	4	5	6	8	9	10

Description: this scale rates your initial impression of the overall appearance of the breasts from an aesthetic point of view.

The scale of the initial impression is based on a rating from 0 (a very poor impression of the overall appearance of the breasts) to 10 (an excellent impression of the overall appearance of the breasts).

If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.2 Symmetry ratings

6.3.2.1 Symmetry of size of breast mounds

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the size of the breast mounds, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the size of one breast mound is severely asymmetrical to the size of the opposite breast mound) to 10 (the sizes of the breast mounds are completely symmetrical).

To rate the symmetry of the size of the left and right breast mounds, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.2 Symmetry of shape of breast mounds

	Severely Asymmetrical									Completely Symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the shape of the breast mounds, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the shape of one breast mound is severely asymmetrical to the shape of the opposite breast mound) to 10 (the shapes of the breast mounds are completely symmetrical).

To rate the symmetry of the shape of the left and right breast mounds, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.3 Symmetry of level - lowest visible point of breasts

	Severely Asymmetrical									Completely Symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the lowest visible point of one breast to the lowest visible point of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the lowest visible point of one breast is severely asymmetrical to the lowest visible point of the opposite breast) to 10 (the lowest visible points of the breasts are completely symmetrical).

To rate the symmetry of the lowest visible point of the left and right breasts, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.4 Symmetry of skin color of breasts

	Severely Asymmetrical									Completely Symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the skin color of the breast mounds, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the skin color of one breast is severely asymmetrical to the skin color of the opposite breast) to 10 (the skin color of the breasts is completely symmetrical). An area of skin with markedly different color would indicate poor symmetry.

To rate the symmetry of the skin color of the left and right breast, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.5 Symmetry of nipple/areola complex position (NAC)

	Severely Asymmetrical									Completely Symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the position of the nipple/areola complex (NAC) relative to the position of the NAC of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (The position of the NAC on one breast is severely asymmetrical to the position of the NAC of the opposite breast) to 10 (The position of the NAC on one breast is completely symmetrical to the position of the NAC of the opposite breast).

To rate the symmetry of NAC relative to midline, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.6 Symmetry of nipple position

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the position of the nipple of one breast to the position of the nipple of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the position of the nipple of one breast is severely asymmetrical to the position of the nipple of the opposite breast) to 10 (the position of the nipple is completely symmetrical to the position of the nipple of the opposite breast).

To rate the symmetry of position of the left and right nipples, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.7 Symmetry of areola position

	Severely Asymmetrical									Completely Symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the position of the areola of one breast relative to the position of the areola of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (The position of the areola of one breast is severely asymmetrical to the position of the areola of the opposite breast) to 10 (The position of the areola of one breast is completely symmetrical to the position of the areola of the opposite breast).

To rate the symmetry of areolar positions, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.8 Symmetry of nipple diameter

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the diameter of the nipple of one breast to the diameter of the nipple of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the diameter of the nipple of one breast is severely asymmetrical to the diameter of the nipple of the opposite breast) to 10 (the diameter of the nipple is completely symmetrical to the diameter of the nipple of the opposite breast).

To rate the symmetry of diameters of the left and right nipples, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.9 Symmetry of areola diameter

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the diameter of the areola of one breast relative to the diameter of the areola of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (The diameter of the areola of one breast is severely asymmetrical to the diameter of the areola of the opposite breast) to 10 (The diameter of the areola of one breast is completely symmetrical to the diameter of the areola of the opposite breast). If the areolae are not exactly circular, consider the average diameters.

To rate the symmetry of areolar diameters, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.10 Symmetry of nipple color

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the color of the nipple of one breast to the color of the nipple of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the color of the nipple of one breast is severely asymmetrical to the color of the nipple of the opposite breast) to 10 (the color of the nipple is completely symmetrical to the color of the nipple of the opposite breast).

To rate the symmetry of color of the left and right nipples, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.2.11 Symmetry of areola color

	Severely asymmetrical									Completely symmetrical
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the symmetry of the color of the areola of one breast to the color of the areola of the contralateral or opposite breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the color of the areola of one breast is severely asymmetrical to the color of the areola of the opposite breast) to 10 (the color of the areola is completely symmetrical to the color of the areola of the opposite breast).

To rate the symmetry of color of the left and right areolae, select a rating from 0 to 10. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate symmetry because left breast is missing.

6.3.3 Individual ratings

6.3.3.1 Overall breast size relative to the visible torso

	Too Small					Just Right					Too Large
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale rates your impression of the overall size of each breast relative to the size of the patient's visible torso (trunk).

The scale of the overall breast size is based on a rating from 0 (breast is too small relative to the visible torso) to 10 (the breast appears too large relative to the visible torso) with a rating of 5 equal to the breast size is just right relative to the visible torso.

This item is rated separately for the left and right breasts. If you feel this item cannot be properly rated, select "N/A". You will be prompted to explain this selection, e.g., cannot rate it because left breast is missing.

6.3.3.2 Ptosis and Pseudoptosis

Description: the scale is the current system that is accepted and used by most surgeons to rate the degree of breast droopiness or sagging from the chest wall. The ptosis scale is based on the location of the nipple relative to the level of the inframammary fold (IMF) while the presence of absence of pseudoptosis relates to whether the breast tissue is located above or below the level of the IMF with the nipple above the IMF. The figures on the following pages give a visible and written definition for the extent of ptosis and pseudoptosis.

Ptosis

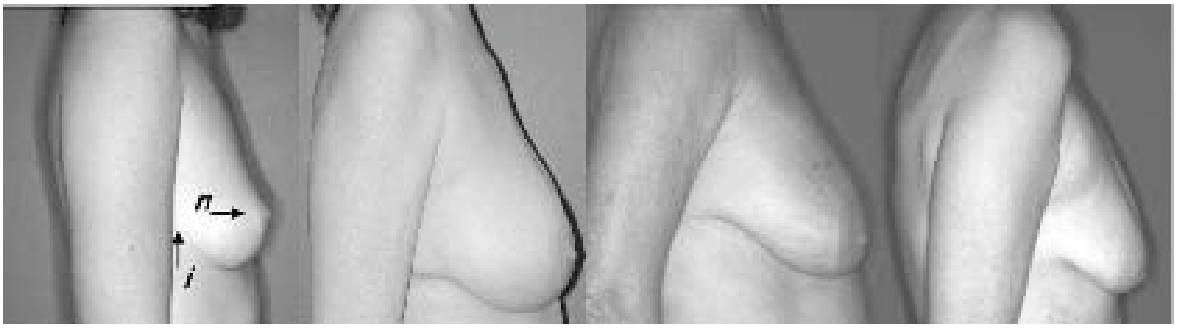
Ptosis is rated separately for each breast. To rate the extent of ptosis for a breast, select a rating from 0 to 3. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

	No Ptosis	Minor Ptosis	Moderate Ptosis	Major Ptosis
N/A	0	1	2	3

Pseudoptosis

Pseudoptosis is rated separately for each breast. To rate the presence or absence of pseudoptosis, select a rating of Y (yes) or N (no). If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

N/A	No	Yes



Grade 0 Grade 1 Grade 2 Grade 3

Figure 6.2: Sample images (right lateral view) illustrating the ptosis grades described by Regnault [16] and Bostwick [15]. *n*: nipple point, *i*: lateral terminus of inframammary fold.

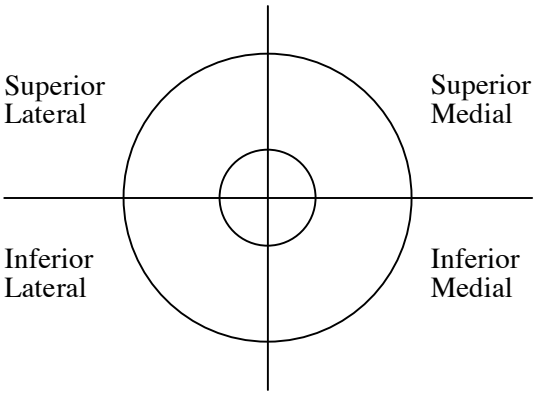
6.3.3.3 Isolated contour deformity

Description: this scale rates the severity of a contour deformity by its location in a specific section on each breast. A contour deformity is a focal or isolated area where there is a deviation of the expected contour for one or more sections of the breast. These contour deviations include a focal area of excessive tissue, a dent, a depression or some combination of these characteristics.

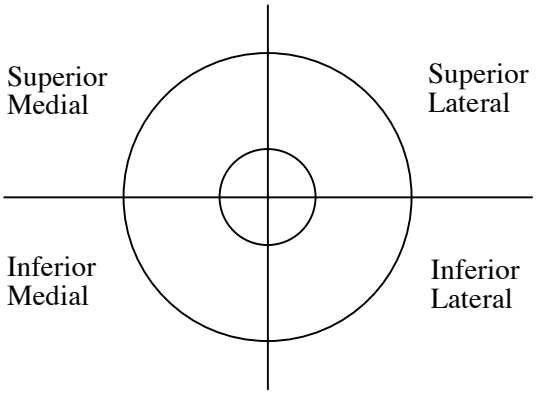
The scale is rated from 0 (there is a marked contour deformity) to 10 (there is no contour deformity). A value is selected for each of Superior Medial, Superior Lateral, Inferior Medial, and Inferior lateral regions of each breast. The quadrants are defined relative to the nipple areolar complex as shown below.

If you feel this item cannot be properly rated, select "N/A". You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

	Marked Contour Deformity									No Contour Deformity
N/A	0	1	2	3	4	5	6	7	8	9	10



Right Breast



Left Breast

Definition of quadrants of the breasts: The larger outer circle represents the breast. The smaller inner circle represents the areola with the crosshairs centered at the nipple.

6.3.3.4 Contour of breasts

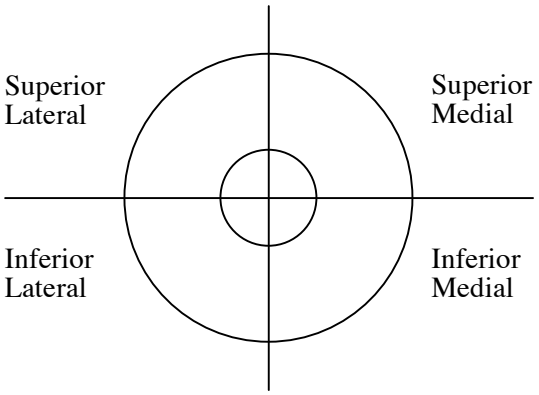
Description: this scale rates the deviation of the contour relative to the perfect or ideal contour for a specific section on each breast.

The scale is rated from 0 (the section is markedly deficient in volume relative to the perfect volume contour for that section) to 10 the section is overly full in volume relative to the perfect volume contour for that section) with 5 (the section has perfect contour).

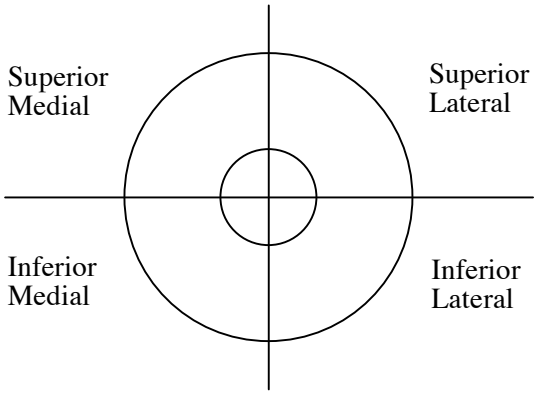
A value is selected for each of Superior Medial, Superior Lateral, Inferior Medial, and Inferior lateral regions of each breast. The quadrants are defined relative to the nipple areolar complex.

If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

	Markedly Deficient					Perfect				Overly Full
N/A	0	1	2	3	4	5	6	7	8	9	10



Right Breast



Left Breast

Definition of quadrants of the breasts: The larger outer circle represents the breast. The smaller inner circle represents the areola with the crosshairs centered at the nipple.

6.3.3.5 Natural shape

	Very unnatural shape									Very natural shape
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale rates your impression of the overall natural appearance of the shape of each breast.

The scale of the naturally shaped breast is based on a rating from 0 (a very unnatural shape of the breast) to 10 (a very natural shape of the breast).

A natural shaped breast should be compatible with the native breast. The breast should not look surgically altered nor have the appearance of a breast augmentation, reduction, or mastopexy. Some characteristics of an unnaturally shaped breast are superior pole fullness, elements of a distorted contour, and excessively round contours.

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.6 Scars - Overall appearance relative to surrounding skin

	Very obvious, noticeable									No visible scars
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the overall appearance of the scars of a breast to the skin of that breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the overall appearance of the scars is very obvious and noticeable relative to the skin of the breast) to 10 (there are no visible scars on the breast).

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.7 Scars - Thickness relative to surrounding skin

	Very thick scar					Flat scar				Depressed scar	
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the thickness (vertical height) of the scars of a breast to the level of the skin of that breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the scar is a very thick scar compared to the level of the skin of the breast) to 10 (the scar is lower in vertical height compared to the level of the breast skin).

Scar thickness refers to the vertical height of the scar relative to surrounding skin (e.g., a hypertropic scar would be a very thick scar).

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.8 Scars - Width relative to surrounding skin

	Very wide scar					Moderate				Thin scar
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: Some scars vary in width along their course, but this scale is used to rate the overall width of the scars on the breast, i.e., the overall spread of the scar on the breast skin.

This scale compares the overall width of the scars of a breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the scars is a very wide scar on the skin of the breast) to 10 (the scar is very thin and barely noticeable on the breast skin).

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.9 Scars - Color relative to surrounding skin

	Very different color from surrounding skin									Very similar color from surrounding skin
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the color of the scars of a breast to the color of the skin of that breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the color of the scars is a very different color from the color of the skin around the breast) to 10 (the color of the scars is a very similar color from the color of the skin around the breast).

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.10 Diameter of nipple relative to breast size

	Too small					Just right					Too large
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the diameter of the nipple of a breast to the size of the breast, itself.

The scale is based on a rating from 0 (the diameter of the nipple is much too small relative to the size of the breast) to 10 (the diameter of the areola is much too large relative to the size of the breast) with a rating of 5 indicating that the diameter of the nipple relative to the breast size is just right.

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.11 Diameter of areola relative to breast size

	Too small					Just right					Too large
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the diameter of the areola of a breast to the size of the breast, itself.

The scale is based on a rating from 0 (the diameter of the areola is much too small compared to the size of the breast) to 10 (the diameter of the areola is much too large compared to the size of the breast) with a rating of 5 indicating that the diameter of the areola relative to the breast size is just right.

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.12 Coloration of nipple relative to color of areola

	Much too light					Just right				Much too dark
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the color of the nipple of a breast to the color of the areola around the nipple of that breast, irrespective of the composition of the skin of the nipple and/or the areola, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the color of the nipple is much too light compared to the color of the areola of the breast) to 10 (the color of the nipple is much too dark compared to the color of the areola of the breast) with 5 (the color of the nipple is just right compared to the color of the areola of the breast).

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.13 Coloration of areola relative to surrounding skin

	Much too light					Just right				Much too dark
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale compares the color of the areola of a breast to the color of the skin around the areola of that breast, irrespective of the composition of the skin of the breast, i.e., native breast skin, flap skin, or some combination of the two types of skin.

The scale is based on a rating from 0 (the color of the areola is much too light compared to the color of the skin around that areola of the breast) to 10 (the color of the areola is much too dark compared to the color of the skin around that areola of the breast) and 5 (the color of the areola is just right compared to the color of the skin around the areola.

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

6.3.3.14 Projection of nipple relative to surrounding skin

None						Just right					Excessive
N/A	0	1	2	3	4	5	6	7	8	9	10

Description: this scale rates the projection of the nipple of a breast relative to the level of the skin around the nipple of that breast.

The scale is based on a rating from 0 (there is no projection of the nipple relative to the level of the skin around that nipple of the breast) to 10 (the projection of the nipple is too excessive relative to the level of the skin around that nipple of the breast) with a value of 5 indicating that the nipple projection is just right relative to the level of the surrounding skin.

This item is rated separately for each breast. If you feel this item cannot be properly rated, select “N/A”. You will be prompted to explain this selection, e.g., cannot rate item because left breast is missing.

Chapter 7: UNDERSTANDING SURGEONS' ASSESSMENTS OF THE AESTHETIC OUTCOME OF BREAST CANCER TREATMENT USING EYE-TRACKING

Contribution and publication: the study in this chapter has been submitted for consideration of presentation at a conference on medical image perception.

7.1 INTRODUCTION

As described in Chapter 6, we are developing an observer rating scale under the leadership of an experienced behavioral scientist Dr. Basen-Engquist (Behavioral Science, UT MDACC) to enable plastic surgeons to subjectively, but reliably, describe breast anatomy. During the discussions to generate items for the breast aesthetics observer rating scale, it became apparent that it is challenging for observers to verbalize what they are looking for or even to be conscious of how they assess aesthetic properties. The goal of this study was to understand how plastic surgeons assess breast aesthetics by using eye-tracking technology to record their gaze path while they rate breast anatomy on clinical photographs (Figure 7.1).

To gain an understanding of what the eye is naturally attracted to, studies examining visual attention on image were reviewed [196]. In recent models, visual attention is broken down into localized analysis problems. Neurons at the earliest stages of processing are tuned to visual attributes including contrast, intensity, color contrast, orientation, direction, spatial frequency, and velocity of motion. Because recognition and identification of visual stimuli take place at the same time as visual attention (though in different parts of the brain), it is assumed that the surgeon in the experiment can quickly recognize abnormalities and focus his/her attention on those features instead of features that would draw the attention of a naive observer (contrast, intensity, etc.).

Figure 7.1: The image shows an example of desktop eye-tracking equipment from Applied Science Laboratories (Bedford, MA), which can unobtrusively collect eye-position data while a subject examines an image presented on the computer screen.

Eye-tracking has been used to help understand medical image perception. Numerous of studies have been conducted on the visual search processes of radiologists. Among these include an experiment studying the affect of lesion conspicuity on the visual search strategy of radiologists' in mammogram reading [197], one studying visual scanning patterns and the relationship between duration of gaze at lesion and correct diagnosis of it [198], one comparing the relationship between lesion subtlety and detection of the lesion [199], and another study using an artificial neural network to predict radiologists' diagnosis using spatial frequency representation of regions [200]. All these studies focus on understanding the visual search pattern and the resulting accuracy in mammogram. By comparison, I am unaware of any previous studies in plastic surgery that employed eye-tracking technology were unexplored. Thus, there is a clear opportunity to introduce this powerful methodology to a new clinical research area.

7.2 MATERIALS AND METHODS

7.2.1 Eye-tracker

An Applied Science Laboratories (Bedford, MA) Eye-Tracking System (Model 504) was used to track the gaze of plastic surgeons at the Behavioral Research and Treatment Center at The University of Texas M. D. Anderson Cancer Center (UT MDACC) (Figure 7.1). I was trained using identical equipment in the LIVE Laboratory at UT Austin. The research coordinator based at UT MDACC (Angela Burgess) and I conducted the eye tracking experiments at UT MDACC. The Model 504 eye tracker is an optional magnetic head tracker where the eye is illuminated by the beam from near infrared LED's on the pan/tilt optics module. An auto-focusing lens system in the pan/tilt module focuses a telephoto image of the eye onto a video sensor (eye-camera). The pan/tilt mechanism can rotate the illumination source in order to follow the eye as the subject moves his/her gaze. According to the specifications provided by the eye-tracker manufacturer, the measurable field of view is about 25 degrees visual angle to either side of the optics, about 25 degrees above the optics and about 10 degrees below the optics. Precision refers to how closely individual measurements agree with each other. The precision of the Model 504 eye tracker is better than 0.5 degrees of the visual angle, which corresponds to a circle of $0.55\,cm$ (1/4 inch) in radius on the monitor which displays the image at a viewing distance of 63.5 cm (25 inch). Accuracy refers to how closely a measured value agrees with the correct value. The accuracy between the true eye position and the computed measurement is less than 1 degree of the visual angle $1.11\,cm$ (1/2 inch), which corresponds to a circle of $0.11\,cm$ in radius on the monitor under the viewing conditions defined above (Figure 7.2).

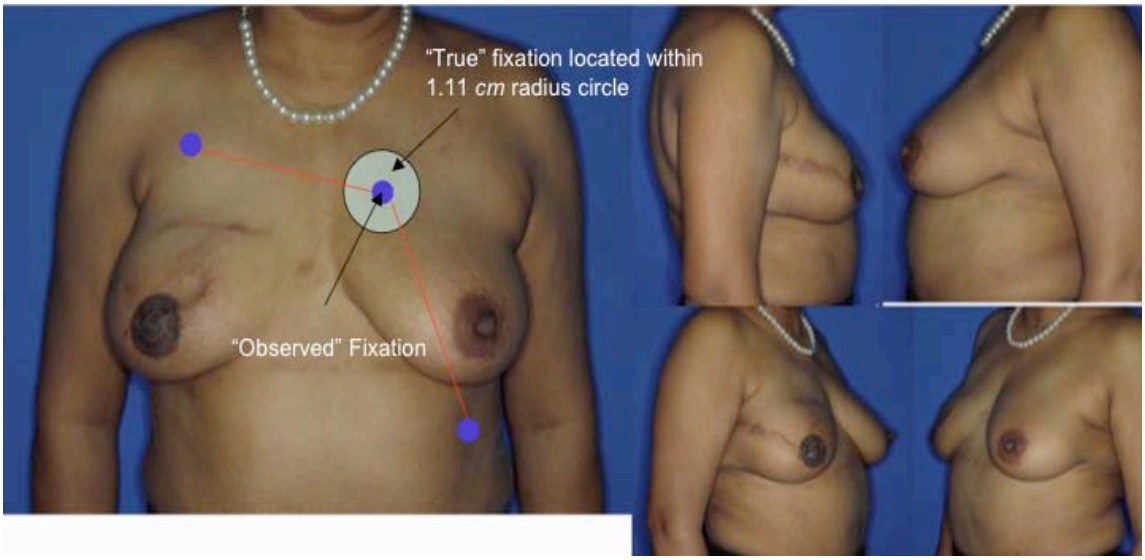


Figure 7.2: This image illustrates the accuracy of the eye-tracker ASL 540. This model has a accuracy of 1 degree in visual angle which is represented as a circle with 1.11 *cm* (½ inch) in radius.

7.2.2 Calibration

In the setup at the UT MDACC laboratory, the system was connected to a computer that ran the calibration (Figure 7.3) and interface software (ASL) and saved the data, while another computer was used to display the images in a darkened room. In this setup, the subject was seated in the eye-tracker room equipped with monitoring facilities that allows communication with monitoring room. Images were displayed on a standard LCD monitor with 1024 by 768 pixels resolution using the Internet Explorer (Microsoft) browser. The raw data measured by the eye-tracker are the separation between the pupil center and the corneal reflection (CR). The relation between these raw values and the eye line of gaze differs for each subject and for different optical unit. The purpose of the calibration process is to provide data that will allow the eye-tracker to account for individual subject differences.

The objective is to have the subject look at (fixate on) each of the nine calibration points, which are at known locations. This procedure must be performed for every subject. The points are numbered from left to right; 1-3 for the top row, 4-6 for the middle row, and 7-9 for the bottom row. The actual distribution of the nine points is taken from the scene monitor and entered into memory with the eye-tracker "set target points" function. The points cover about 80 percent of the monitor screen area and are separated by 15-20 degrees visual angle horizontally and 10-15 degrees vertically. These are ideal specifications. The system was calibrated at the beginning of the session and repeated before each case.

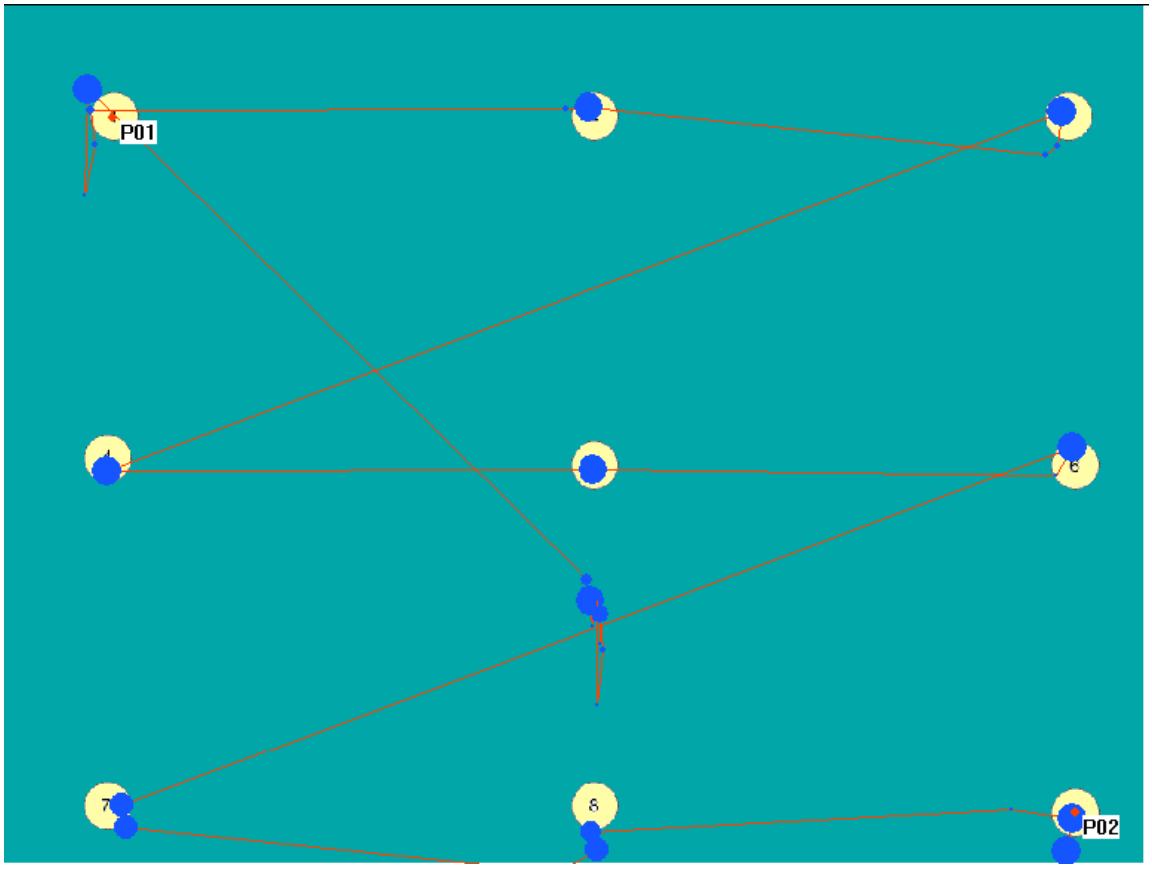


Figure 7.3: This image illustrates the calibration map image provided by the ASL eye-tracker manufacture. The 9 points cover about 80 percent of the monitor screen area and are separated by 15-20 degrees visual angle horizontally and 10-15 degrees vertically. All points are numbered from left to right; 1-3 for the top row, 4-6 for the middle row, and 7-9 for the bottom row. The blue dots indicate the gaze of the observer recorded during the calibration procedure. The observer was directed to start looking at from number 1and to proceed sequentially to number 9. Recalibration was performed as necessary.

7.2.3 Data collection

Eye-data were collected from two plastic surgeons who viewed clinical photographs of 8 women who underwent breast reconstruction. A web-based interface was used to display the clinical images and collect observer ratings of aesthetics (Chapter 6) during the eye-tracker experiments (Figure 7.4). Five images of each patient are

displayed: an anterior-posterior and lateral and oblique views of both the right and left breasts. The oblique views are taken at an angle of about 45 degrees from the true lateral. The vertical extent of the views is from just below the chin down to the top of the pubic bone. The oblique views are taken at an angle of about 45 degrees from the true lateral. All images were displayed on a standard LCD monitor with 1024 by 768 pixels resolution using the Internet Explorer (© Microsoft Corporation) browser.

Two plastic surgeons rated breast aesthetics using the top-six most influential rating items selected from a preliminary breast aesthetic rating instrument developed under the leadership of an experienced behavioral scientist Dr. Basen-Engquist (Behavioral Science, UT MDACC). The rating items used were: (1) initial impression of overall appearance of the breasts, (2) symmetry of size of breast mounds, (3) symmetry of shape of breast mounds, (4) aesthetic shape, (5) natural shape, and (6) final impression of overall appearance of the breasts (Refer to Chapter 6 for details on the rating items). A specific rating item is displayed before and after the images are displayed, but not while they are displayed to prevent the observer from being distracted by the rating item text. Vertical and horizontal coordinates of eye position and pupil diameter were saved to an "Eyedat file" on the Interface PC hard disk. In addition, event marks were entered from the keyboard to separate the recordings for each item. A field of data, consisting of the elements just listed, is recorded every 60th of a second (60 Hz update rate).

7.2.4 Data analysis

EyeNal and Fixplot (ASL, Bedford, MA) were used to plot and analyze the raw eye data on the images (Figure 7.4). Fixation is defined as the mean eye position over a minimum time period (0.1 sec) during which the gaze stays within 1 visual angle. Dwell

is defined as the amount of time during which a contiguous series of 1 or more fixations remains within a regions of interest (ROI). In this chapter, mean dwell times across rating items, patients, and observers were compared. I used the dwell time, rather than fixation time, because the latter is subject to more noise. The relationships between the eye-position and the surgeons' observer ratings were also analyzed.

Regions of interests (ROIs) were used to further analyze the data. Two sets of regions were created, one for regions around each of the 5 photographic views (five ROIs), and another for regions around breast within the 5 photographic views (six ROIs): (1) left breast in AP view, (2) right breast in AP view, (3) both breasts in the left oblique view, (4) both breasts in the right oblique view, (5) left breast in the left lateral view, and (6) right breast in the right lateral view. Fixations were then plotted for each individual item with the regions of interest. Dwell-time tables and transition tables were analyzed.

7.3 RESULTS

7.3.1 Dwell time analysis

Dwell time across the five views (an anterior-posterior and lateral and oblique views of both the right and left breasts) was analyzed and mean dwell time across rating items, patients, and observers were compared (Figure 7.4, Table 7.1). The average time that observer 1 (14.00 seconds) required to assess a case was about one and half times that of observer 2 (19.94 seconds), demonstrating that there is some inter-observer variability in how plastic surgeons approach this task (Table 7.1). Observer 1 spent more time on the items pertaining to shape than to the other items, whereas observer 2 spent the most time on the first item (initial impression of overall appearance) and less time on subsequent items. However, some consistent patterns in dwell time were also seen across

observers and cases. For example, both surgeons spent more time looking at the AP views than at the lateral and oblique views (Observer 1 54%, Observer 2 72%).

Dwell time across the breast regions (separate regions in anterior-posterior and lateral and oblique views of both the right and left breasts) was analyzed and mean dwell time across rating items, patients, and observers were compared (Figure 7.5, Table 7.2). Results similar to those repeated in Table 7.1 were observed on time spent either across views or across items compared to the result in Table 7.1. When the dwell time was compared across the regions of breast, the results showed that both surgeons spent the most time on the breast regions in AP views, specifically more time on left breast regions than right breast region. Overall time spent across the patient cases reduced for both surgeons (O1: 94 sec, O2: 2min 82 sec). This can be explained by that the amount of dwells on other fiudcial points such as sternal notch or navel were removed as only breast regions were considered.

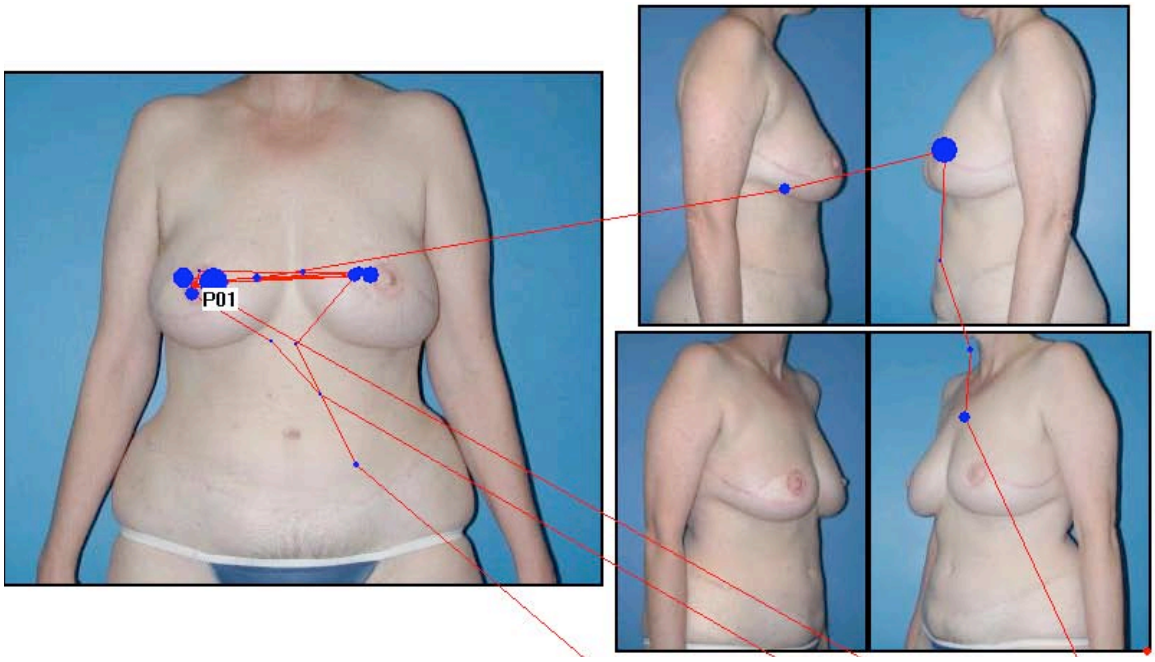


Figure 7.4: This image illustrates an example of a visual scan path recorded by eye-tracking equipment while an observer examined a case to rate the final impression of the overall appearance of the breasts. The observer rated this case as 9 on the 11-point scale, where 0 indicates very poor and 10 indicates excellent. It is apparent that the observer's eyes dwelled longer on the AP view rather than on other views. The scan path also demonstrates that the observer's eyes were drawn to the breast regions where surgical scars were present. Five images are displayed from each patient: an anterior-posterior (AP) and lateral and oblique views of both the right and left breasts. A specific item is displayed only before and after the images are displayed, but not while they are displayed to prevent the observer from being distracted by the item text.

Table 7.1: Mean dwell time across five views spent by two experienced plastic surgeons for assessing the outcomes of 8 patients who underwent breast reconstruction. When dwell time was compared across the rating items, the results showed that observer 1 spent more time on the items pertaining to shape than on the other items, while observer 2 spent the most time on the first item (initial impression of overall impression) and spent less time on the subsequent items. When the dwell time was compared across the views, the results showed that both surgeons spent the most time on the AP views. Overall, observer 2 spent more time evaluating a case than observer 1 did. The rating items are: (Item 1) initial impression of overall appearance of the breasts, (Item 2) symmetry of size of breast mounds, (Item 3) symmetry of shape of breast mounds, (Item 4) aesthetic shape, (Item 5) natural shape, and (Item 6) final impression of overall appearance of the breasts.

	Item 1		Item 2		Item 3		Item 4		Item 5		Item 6		Overall Time		Percent Time (%)	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
AP	1.84	2.74	1.56	2.16	1.76	1.58	1.88	2.15	1.68	1.02	1.34	1.19	10.07	10.84	72	54
RL	0.15	0.89	0.09	0.39	0.29	0.28	0.20	0.39	0.35	0.42	0.08	0.25	1.17	2.61	8	13
LL	0.05	0.22	0.08	0.22	0.14	0.57	0.19	0.30	0.12	0.43	0.11	0.05	0.69	1.78	5	9
RO	0.21	0.78	0.18	0.38	0.26	0.41	0.12	0.52	0.24	0.09	0.10	0.15	1.12	2.34	8	12
LO	0.02	0.52	0.02	0.41	0.19	0.46	0.25	0.33	0.27	0.29	0.20	0.35	0.95	2.37	7	12
Overall Time	2.27	5.16	1.94	3.56	2.64	3.30	2.65	3.68	2.67	2.24	1.83	1.99	14.00	19.94		
Percent Time (%)	16	26	14	18	19	17	19	18	19	11	13	10				

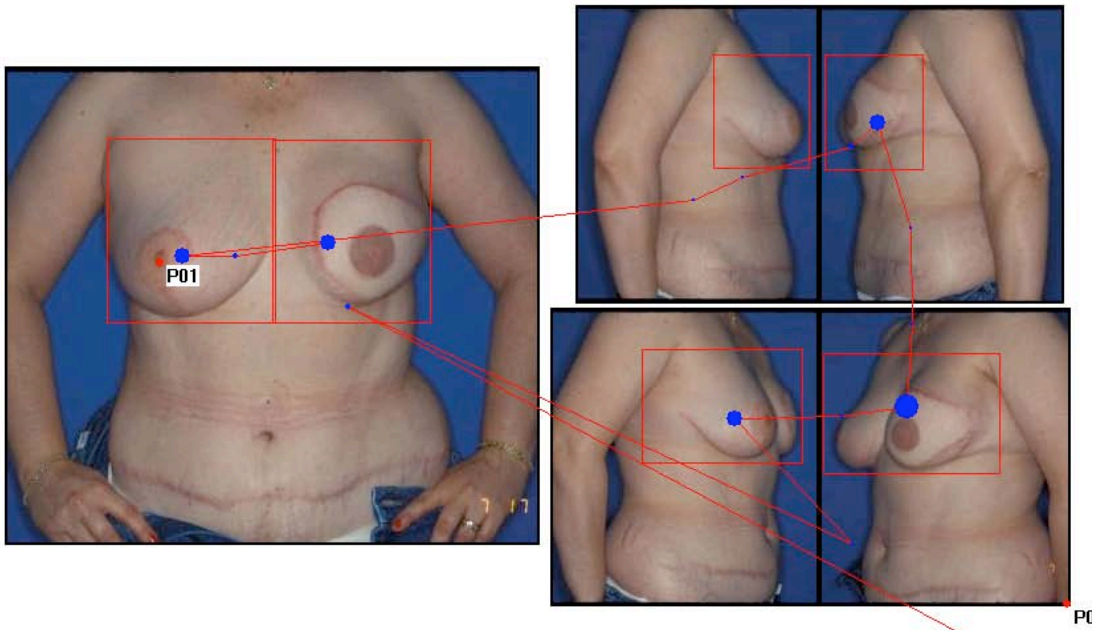


Figure 7.5: This image illustrates an example of a visual scan path recorded by eye-tracking equipment while an observer 2 examined a case to rate the symmetry of size of breast mounds. Regions of interest were created around breast areas in each view. The observer rated this case as 7 on the 11-point scale, where 0 indicates very asymmetric and 10 indicates very symmetric. The scan path demonstrates that the observer's eyes were more drawn to the breast regions where surgical scars were present. Five images are displayed from each patient: an anterior-posterior (AP) and lateral and oblique views of both the right and left breasts. A specific breast aesthetics-related item is displayed only before and after the images are displayed to prevent the observer from being distracted by the item text.

Table 7.2: Mean dwell time across breast regions spent by two experienced plastic surgeons for assessing the outcomes of 8 patients who underwent breast reconstruction. Results similar to those repeated in Table 7.1 were observed. When the dwell time was compared across the regions of breast, the results showed that both surgeons spent the most time on the breast regions in AP views, specifically more time on left breast regions than right breast region.

	Item 1		Item 2		Item 3		Item 4		Item 5		Item 6		Overall Time		Percent Time (%)	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.89	1.18	0.70	0.88	0.68	0.41	1.09	0.82	0.87	0.75	0.31	0.23	4.54	4.27	30	25
APLB	1.13	1.37	0.93	1.26	0.85	0.98	1.41	1.24	0.87	0.26	0.90	0.80	6.10	5.92	41	35
RLB	0.35	0.44	0.21	0.20	0.49	0.22	0.54	0.27	0.45	0.28	0.08	0.00	2.13	1.40	14	8
LLB	0.00	0.14	0.10	0.18	0.12	0.35	0.21	0.21	0.15	0.38	0.10	0.00	0.68	1.27	5	7
ROB	0.22	0.70	0.00	0.37	0.07	0.46	0.19	0.45	0.11	0.05	0.08	0.14	0.67	2.17	4	13
LOB	0.00	0.36	0.07	0.34	0.13	0.49	0.17	0.28	0.25	0.35	0.20	0.28	0.82	2.10	5	12
Overall Time	2.59	4.18	2.01	3.23	2.34	2.91	3.61	3.28	2.71	2.07	1.67	1.46	13.03	17.12		
Percent Time (%)	17	24	13	19	16	17	24	19	18	12	11	9				

7.3.1 Transitions frequency analysis

A transition from region ***i*** to region ***j*** is defined as a dwell period in area ***i*** followed immediately by a dwell period in area ***j***. I analyzed the frequency of transitions between breast regions in each view rather than broad regions of interest around each image view. The regions of interest used were: (APRB) patient's right breast area in AP view, (APLB) patient's left breast area in AP view, (ROB) patient breast area in right oblique view, (LOB) patient's breast area in left oblique view, (RLB) patient's right breast in right lateral view, (LLB) patient's left breast in left lateral view (Figure 7.5). The average number of transitions between regions across the eight cases was computed for each observer for each item. In the transition martrices in Tables 7.3 – 7.8, the region given in the row label is the starting position and the region named in the column label is the ending position.

When transitions between the regions were analyzed, some consistent patterns were seen across observers and rating items. The results showed that there were many high transitions between the breast regions in the AP view (APRB, APLB). The largest number of transitions was between regions APLB to region APRB for the both surgeons across the rating items. This trend was observed regardless of the type of rating items: both rating items for comparing symmetry between breasts (Table 7.4, Table 7.5) and individual ratings for the individual breast (Table 7.6, Table 7.7). Few transitions were observed between the oblique views, lateral views, or combinations of two. This result is consistent with our expectations that the surgeons primarily use the breast regions in AP views and move on to either oblique or lateral views only if needed.

Table 7.3: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the (Item 1) initial impression of overall appearance of the breasts. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for the both surgeons, O1 (1.5) and O2 (1.3).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.9	0.9	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0
APLB	1.5	1.3	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0
ROB	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.0	0.0
LOB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1
RLB	0.0	0.0	0.3	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.3
LLB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0

Table 7.4: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the (Item 2) symmetry of size of breast mounds. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for the both surgeons, O1 (1.0), and O2 (1.3).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APLB	1.0	1.3	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ROB	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.3	0.1	0.0	0.0	0.0
LOB	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
RLB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1
LLB	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0

Table 7.5: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the (Item 3) symmetry of shape of breast mounds. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for the both surgeons, O1 (1.0), and O2 (0.5).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.9	0.1	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APLB	1.1	0.5	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ROB	0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.4	0.4	0.1	0.0	0.0
LOB	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
RLB	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.3
LLB	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0

Table 7.6: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the (Item 4) aesthetic shape. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for the both surgeons, O1 (1.0), and O2 (1.0).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.9	0.7	0.4	0.0	0.0	0.0	0.0	0.0	0.1	0.0
APLB	1.0	1.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0
ROB	0.1	0.1	0.0	0.0	0.0	0.0	0.3	0.1	0.3	0.0	0.0	0.0
LOB	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
RLB	0.1	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
LLB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0

Table 7.7: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the (Item 5) natural shape. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for surgeon, O1 (1.1), and from right breast area to left in oblique and lateral views for surgeon O2 (0.4).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.5	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APLB	1.1	0.2	0.0	0.0	0.1	0.2	0.1	0.0	0.0	0.2	0.1	0.0
ROB	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.3	0.0	0.1	0.0
LOB	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.3	0.0
RLB	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4
LLB	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0

Table 7.8: Mean transition activity across six regions of interest recorded, while an observer (GPR) examined 8 patients who underwent breast reconstruction case to rate the Q6 (Item 6) final impression of overall appearance of the breasts. When transition was compared between the regions, the results showed that there were notably high transitions were observed pertaining breast regions in AP views (APRB, APLB). The highest transitional activity was from region APLB to region APRB for surgeon, O1 (0.9), and from region APLB to region APRB for surgeon, O2 (0.5).

	APRB		APLB		ROB		LOB		RLB		LLB	
	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2	O1	O2
APRB	0.0	0.0	0.4	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APLB	0.9	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0
ROB	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
LOB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
RLB	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
LLB	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.3	0.5	0.0	0.0

7.4 DISCUSSION AND CONCLUSION

The goal of the study was to understand how plastic surgeons assess breast aesthetics by using eye-tracking technology to record their gaze path while they rate breast anatomy on clinical photographs. When the dwell time was compared across the views, the results showed that both surgeons spent the most time on the AP views. Similarly, when transition activity between regions was analyzed, there were many transitions observed between the breast regions in the AP view, while few transitions were observed between other views. This result is consistent with our expectations that the surgeons primarily use the breast regions in the AP view and move on to either oblique or lateral views if needed. One possible explanation for this trend of a great amount of time on and transitions between the breast regions in the AP view is that the eye is searching for asymmetry between the left and right breasts.

There are key limitations of this study on using eye-tracking for the assessment of aesthetic outcomes. First, even though extreme care was taken to control the experimental conditions, some inevitable variability was introduced because of uncontrollable variables such as technical difficulties and the physical condition of the subjects. The surgeons who participated in this study suffered from extreme fatigue that resulted in longer calibration time. Second, because of the limited number of observers and cases in this study, further investigations are needed. It is critical to repeat this study with different surgeons who were not aware of the patient cases used in this study to compare the results.

Chapter 8: DISCUSSION AND CONCLUSION

8.1 SUMMARY OF WORK

Our long-term goal is to develop decision aids that will improve breast cancer treatment. Quantitative, objective measures of aesthetic outcomes with high reliability are needed to meaningfully relate patient and surgical variables to aesthetic outcomes and to compare the outcomes of different breast cancer treatment strategies.

In Chapter 2, I briefly review the literature on the various breast aesthetics outcome assessment methods currently used: subjective ratings, physical measurements, photographic measurements, and 3-D measurements. In Chapter 3, innovative quantitative, objective measurements of breast ptosis based on ratios of distances between fiducial points manually identified in lateral and oblique views of clinical photographs were proposed. I found that an existing subjective scale for rating ptosis showed high intra-observer agreement, but lower inter-observer agreement. New objective measures of ptosis showed encouraging levels of concordance with ratings made by experienced surgeons using the subjective scale. The objective measures were found to be robust to intra- and inter- observer variability in marking fiducial points, even when annotations by "novice" observers were used. In Chapter 4, I compared color measurements obtained by clinical digital photography to those from a standard colorimeter. I found that photography can be used in place of a colorimeter for measuring color properties of skin in a controlled setting. To investigate the validity of using photography for the assessment of scars, in Chapter 5, I presented results on using clinical photographs for the assessment of linear breast surgical scars. Digital clinical photographs were used to assess quantitative, objective measurements of breast surgical scars based on complete color intensity image analysis and area measurements. I demonstrated that the new

objective measures utilizing clinical photographs were effective. In Chapter 6, an observer rating scale of 11 symmetry ratings items, 14 individual breast ratings items, and a global rating on overall appearance before and after the entire rating items was proposed. The observer rating scale was assessed for intra- and inter-observer variability using rating data collected by reconstructive surgeons not involved in designing the scale. In Chapter 7, eye-data were collected from two plastic surgeons, while they rated breast aesthetics based on clinical photographs of 8 women who underwent breast reconstruction. Mean dwell time and transitions between regions of interest across rating items, patients, and observers were compared. Some consistent patterns in dwell time were seen across observers and cases. I found the results of the eye data recorded while the observer rated aesthetic qualities resembled that of the asymmetry search task investigated in other studies.

8.2 SUGGESTIONS FOR FUTURE STUDIES

As described in Chapter 2, breast cancer survivors face a myriad of choices to be made with the assistance of their multi-disciplinary breast care team. They must make decisions about surgical, radiation, chemotherapy, and endocrine therapy options. They must confront these issues with incomplete and, at times, contradictory information. A system for objective assessment of breast aesthetics is critical in order for breast cancer survivors to make right decisions for different reconstructive procedures.

I have proposed quantitative outcome assessment methods using clinical photographs for two representative aesthetic features: ptosis and scars. Eye-tracking was used to understand how the surgeons assess the outcomes while they are looking at the clinical photographs. Substantial future work is needed, however, to fully quantitatively

understand breast outcomes and provide objective data to clinicians and breast cancer survivors.

First, further development is needed to complete and validate a subjective but explicit observer rating scale to rate full range of breast aesthetics based on patients' opinions as well as the knowledge of experienced reconstructive surgeons is needed. Both the observer rating scale and quantitative methods should be assessed for the intra- and inter-observer variability against each other using rating data collected by reconstructive surgeons and patients not involved in designing the scale.

Second, image processing techniques to automatically locate the fiducial points should be developed. Previous studies have all used manually identified fiducial points by having a surgeon mark them when reviewing the images on a computer or photograph in print. One drawback to this approach is that reconstructive surgeons don't routinely review the images in this manner and one needs to be cautious about developing medical decision aids that are dependent on clinicians performing tasks that aren't a part of their normal workflow. A more fully automated system would be more practical to collect large data sets for outcome studies. Reconstructions using autologous tissue typically take 4-12 hours and the most important decisions relating to breast shaping must be made towards the very end. Thus, an automated system would probably be necessary for any intra-operative analysis, since near real-time performance is needed. The automated system should be evaluated for the sensitivity of the algorithm for computing objective aesthetic measures against manual method for identifying fiducial points to improve the reliability.

Third, 3D imaging should be explored since it has the potential to accurately measure aspects of breast anatomy that cannot be adequately evaluated using 2D imaging, such as curvature, shape, and volume. From one scan from a 3D camera, it is

possible to reconstruct an image of the breasts from any angle, whereas several standard photographs are needed to approximate the 3D nature of the breasts. However, current 3D imaging systems are expensive and require a dedicated room in the clinic. Moreover, there has been little work to date on developing quantitative measures of breast aesthetics from 3D scans. Aesthetic outcome assessment methods used in 2D imaging should be extended to 3D imaging and the results should be compared to those obtained from 2D images.

Finally, increased collaboration with behavioral scientists is needed to understand how patient satisfaction is related to aesthetics. Current studies have assessed the surgical outcomes by simply using the assessments either from clinicians, patients, or combination of two, but not integrating the aspects of behavioral science which can help clinicians understand patients' assessments. In particular, expertise in body image would enhance our understanding of breast aesthetics and quality of life.

Bibliography

[1] "Cancer Facts and Figures 2007," American Cancer Society 2007.

[2] K. K. Hunt, G. L. Robb, and E. A. Strom, *Breast Cancer*. New York: Springer, 2001.

[3] R. J. Bold, "Surgical management of breast cancer: today and tomorrow," *Cancer Biotherapy & Radiopharmaceuticals.*, vol. 17, pp. 1-9, 2002.

[4] ASPS, "2000/2005/2006 National Plastic Surgery Statistics," American Society of Plastic Surgeons, Arlington Heights, IL 2007.

[5] J. Bostwick, III, *Plastic and reconstructive breast surgery*, 2 ed. St. Louis, Missouri: Quality Medical Publishing, Inc., 1999.

[6] S. S. Kroll, "A comparison of factors affecting aesthetic outcomes of TRAM flap breast reconstructions," *Plastic & Reconstructive Surgery*, vol. 96, pp. 860-864, 1995.

[7] S. S. Kroll, "A comparison of outcomes using three different methods of breast reconstruction," *Plastic & Reconstructive Surgery*, vol. 90, pp. 455-462, 1992.

[8] L. I. Tsouskas and I. S. Fentiman, "Breast compliance: a new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer," *Breast Cancer Research & Treatment*, vol. 15, pp. 185-90, 1990.

[9] V. Sacchini, A. Luini, S. Tana, L. Lozza, V. Galimberti, M. Merson, R. Agresti, P. Veronesi, and M. Greco, "Quantitative and qualitative cosmetic evaluation after conservative treatment for breast cancer," *European Journal of Cancer*, vol. 27, pp. 1395-400, 1991.

[10] K. B. Clough, J. C. Cuminet, A. Fitoussi, C. Nos, and V. Mosseri, "Cosmetic sequelae after conservative treatment for breast cancer: Classification and results of surgical correction," *Annals of Plastic Surgery*, vol. 41, pp. 471-481, 1998.

[11] J. R. Harris, M. B. Levene, G. Svensson, and S. Hellman, "Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast," *Int. J. Radiation Oncology Biol. Phys.*, vol. 5, pp. 257-261, 1979.

[12] S. K. Al-Ghazal, R. W. Blamey, J. Stewart, and A. A. Morgan, "The cosmetic outcome in early breast cancer treated with breast conservation," *European Journal of Surgical Oncology*, vol. 25, pp. 566-70, 1999.

[13] D. Clarke, A. Martinez, and R. S. Cox, "Analysis of cosmetic results and complications in patients with stage I and II breast cancer treated by biopsy and irradiation," *International Journal of Radiation Oncology, Biology, Physics*, vol. 9, pp. 1807-13, 1983.

[14] V. Sacchini, A. Luini, R. Agresti, M. Greco, A. Manzari, L. Mariani, R. Zucali, and B. McCormick, "The influence of radiotherapy on cosmetic outcome after breast conservative surgery," *International Journal of Radiation Oncology, Biology, Physics*, vol. 33, pp. 59-64, 1995.

[15] R. Kuzbari, M. Deutinger, B. P. Todoroff, B. Schneider, and G. Freilinger, "Surgical treatment of developmental asymmetry of the breast. Long term results," *Scandinavian Journal of Plastic & Reconstructive Surgery & Hand Surgery*, vol. 27, pp. 203-7, 1993.

[16] C. M. Malata, J. C. Boot, E. T. Bradbury, A. R. Ramli, and D. T. Sharpe, "Congenital breast asymmetry: subjective and objective assessment," *British Journal of Plastic Surgery*, vol. 47, pp. 95-102, 1994.

[17] B. Stark and N. Olivari, "Breast asymmetry: an objective analysis of postoperative results," *European Journal of Plastic Surgery*, vol. 14, pp. 173-176, 1991.

[18] D. J. Smith, Jr., W. E. Palin, Jr., V. Katch, and J. E. Bennett, "Surgical treatment of congenital breast asymmetry," *Annals of Plastic Surgery*, vol. 17, pp. 92-101, 1986.

[19] Y. Brandberg, M. Malm, and L. Blomqvist, "A Prospective and Randomized Study, "SVEA," Comparing Effects of Three Methods for Delayed Breast Reconstruction on Quality of Life, Patient-Defined Problem Areas of Life, and Cosmetic Result [Articles]," *Plastic & Reconstructive Surgery*, vol. 105, pp. 66-74, 2000.

[20] D. F. Veiga, M. S. Neto, Y. Juliano, L. M. Ferreira, and J. L. B. S. Rocha, "Evaluations of the aesthetic results and patient satisfaction with the breast recnstruction," *Annals of Plastic Surgery*, vol. 48, pp. 515-520, 2002.

[21] C. Vrieling, L. Collette, E. Bartelink, J. H. Borger, S. J. Brenninkmeyer, J. C. Horiot, M. Pierart, P. M. Poortmans, H. Struikmans, E. Van der Schueren, J. A. Van Dongen, E. Van Limbergen, and H. Bartelink, "Validation of the methods of cosmetic assessment after breast-conserving therapy in the EORTC "boost versus no boost" trial. EORTC Radiotherapy and Breast Cancer Cooperative Groups. European Organization for Research and Treatment of Cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 45, pp. 667-76, 1999.

[22] J. C. Lowery, E. G. Wilkins, W. M. Kuzon, and J. A. Davis, "Evaluations of aesthetic results in breast reconstruction: an analysis of reliability," *Annals of Plastic Surgery*, vol. 36, pp. 601-6; discussion 607, 1996.

[23] R. D. Pezner, J. A. Lipsett, N. L. Vora, and K. R. Desai, "Limited usefulness of observer-based cosmesis scales employed to evaluate patients treated conservatively for breast cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 11, pp. 1117-9, 1985.

[24] H. Kaija, S. Rauni, I. Jorma, and H. Matti, "Consistency of patient- and doctor-assessed cosmetic outcome after conservative treatment of breast cancer.," *Breast Cancer Research & Treatment.*, vol. 45, pp. 225-8, 1997.

[25] M. E. Taylor, C. A. Perez, K. J. Halverson, R. R. Kuske, G. W. Philpott, D. M. Garcia, J. E. Mortimer, R. J. Myerson, D. Radford, and C. Rush, "Factors influencing cosmetic results after conservation therapy for breast cancer.," *International Journal of Radiation Oncology, Biology, Physics*, vol. 31, pp. 753-64, 1995.

[26] D. R. Christie, M. Y. O'brien, and J. A. Christie, "A comparison of methods of cosmetic assessment in breast conservation treatment," *Breast*, vol. 5, pp. 358-367, 1996.

[27] A. K. Chawla, L. A. Kachnic, A. G. Taghian, A. Niemierko, D. T. Zapton, and S. N. Powell, "Radiotherapy and breast reconstruction: complications and cosmesis with TRAM versus tissue expander/implant.," *Int J Radiat Oncol Biol Phys.*, vol. 54, pp. 520-6, 2002.

[28] K. C. Sneeuw, N. K. Aaronson, J. R. Yarnold, M. Broderick, J. Regan, G. Ross, and A. Goddard, "Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. Comparison of patients' ratings, observers' ratings, and objective assessments," *Radiotherapy & Oncology*, vol. 25, pp. 153-9, 1992.

[29] D. E. Wazer, T. DiPetrillo, R. Schmidt-Ullrich, L. Weld, T. J. Smith, D. J. Marchant, and N. J. Robert, "Factors influencing cosmetic outcome and complication risk after conservative surgery and radiotherapy for early-stage breast carcinoma.," *Journal of Clinical Oncology*, vol. 10, pp. 356-63, 1992.

[30] D. A. Hidalgo, "Aesthetic refinement in breast reconstruction: complete skin-sparing mastectomy with autogenous tissue transfer.," *Plastic & Reconstructive Surgery*, vol. 102, pp. 63-70, 1998.

[31] D. A. Markiewicz, D. J. Schultz, J. A. Haas, E. E. Harris, K. R. Fox, J. H. Glick, and L. J. Solin, "The effects of sequence and type of chemotherapy and radiation therapy on cosmesis and complications after breast conservation therapy.," *International Journal of Radiation Oncology, Biology, Physics*, vol. 35, pp. 661-8, 1996.

[32] G. R. Ray, V. J. Fish, J. B. Marmor, W. Rogoway, P. Kushlan, C. Arnold, R. H. Lee, and F. Marzoni, "Impact of adjuvant chemotherapy on cosmesis and complications in stages I and II carcinoma of the breast treated by biopsy and radiation therapy," *International Journal of Radiation Oncology, Biology, Physics*, vol. 10, pp. 837-41, 1984.

[33] R. Sarin, K. A. Dinshaw, S. K. Shrivastava, V. Sharma, and S. M. Deore, "Therapeutic factors influencing the cosmetic outcome and late complications in the conservative management of early breast cancer.," *International Journal of Radiation Oncology, Biology, Physics*, vol. 27, pp. 285-92, 1993.

[34] M. Noguchi, Y. Saito, H. Nishijima, M. Koyanagi, A. Nonomura, Y. Mizukami, S. Nakamura, T. Michigishi, N. Ohta, and H. Kitagawa, "The psychological and cosmetic aspects of breast conserving therapy compared with radical mastectomy.," *Surgery Today*, vol. 23, pp. 598-602, 1993.

[35] G. R. Ray and V. J. Fish, "Biopsy and definitive radiation therapy in Stage I and II adenocarcinoma of the female breast: analysis of cosmesis and the role of electron beam supplementation.," *International Journal of Radiation Oncology, Biology, Physics*, vol. 9, pp. 813-8, 1983.

[36] B. Pierquin, R. Owen, C. Maylin, Y. Otmezguine, M. Raynal, W. Mueller, and S. Hannoun, "Radical radiation therapy of breast cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 6, pp. 17-24, 1980.

[37] H. Bartelink, F. van Dam, and J. van Dongen, "Psychological effects of breast conserving therapy in comparison with radical mastectomy.," *Int J Radiat Oncol Biol Phys.*, vol. 11, pp. 381-5, 1985.

[38] P. Romestaing, Y. Lehingue, C. Carrie, R. Coquard, X. Montbarbon, J. M. Ardiet, N. Mamelle, and J. P. Gerard, "Role of a 10-Gy boost in the conservative treatment of early breast cancer: results of a randomized clinical trial in Lyon, France.," *Journal of Clinical Oncology*, vol. 15, pp. 963-8, 1997.

[39] R. P. Zimmerman, R. J. Mark, A. I. Kim, T. Walton, D. Sayah, G. F. Juillard, and M. Nguyen, "Radiation tolerance of transverse rectus abdominis myocutaneous-free flaps used in immediate breast reconstruction.," *American Journal of Clinical Oncology*, vol. 21, pp. 381-5, 1998.

[40] M. E. Brown, J. Semple, and P. Neligan, "Variables affecting symmetry of the nipple-areola complex," *Plastic & Reconstructive Surgery*, vol. 96, pp. 846, 1995.

[41] M. Noguchi, M. Minami, M. Earashi, T. Taniya, I. Miyazaki, H. Nishijima, T. Takanaka, H. Kawashima, Y. Saito, and S. Nakamura, "Oncologic and cosmetic outcome in patients with breast cancer treated with wide excision, transposition of adipose tissue with latissimus dorsi muscle, and axillary dissection followed by radiotherapy.," *Breast Cancer Res Treat.*, vol. 35, pp. 163-71, 1995.

[42] V. F. Cocquyta, P. N. Blondeelb, H. T. Depyperec, K. A. Van De Sijpeb, K. K. Daemsd, S. J. Monstreyb, and S. J. P. Van Belle, "Better cosmetic results and comparable quality of life after skin-sparing mastectomy and immediate autologous breast reconstruction compared to breast conservative treatment," *The British Association of Plastic Surgeon*, vol. 56, pp. 462-470, 2003.

[43] A. D. Mandrekas, G. J. Zambacos, A. Anastasopoulos, D. Hapsas, N. Lambrinaki, and L. Ioannidou-Mouzaka, "Aesthetic Reconstruction of the Tuberous Breast Deformity," *Plastic & Reconstructive Surgery*, vol. 112, pp. 1099-1108, 2003.

[44] N. ShaikhNaidu, B. A. Preminger, K. Rogers, P. Messina, and L. B. Gayle, "Determinants of Aesthetic Satisfaction Following TRAM and Implant Breast Reconstruction," *Annals of Plastic Surgery*, vol. 52, pp. 465-470, 2004.

[45] L. A. Woerdeman, J. J. Hage, E. A. Thio, F. A. Zoetmulder, and E. J. Rutgers, "Breast-conserving therapy in patients with a relatively large (T2 or T3) breast cancer: long-term local control and cosmetic outcome of a feasibility study," *Plastic & Reconstructive Surgery.*, vol. 113, pp. 1607-16, 2004.

[46] P. G. Cordeiro, A. L. Pusic, J. J. Disa, B. McCormick, and K. VanZee, "Irradiation after Immediate Tissue Expander/Implant Breast Reconstruction: Outcomes, Complications, Aesthetic Results, and Satisfaction among 156 patients," *Plastic and Reconstructive Surgery*, vol. 113, pp. 877-881, 2003.

[47] D. J. Smith, Jr., W. E. Palin, Jr., V. L. Katch, and J. E. Bennett, "Breast volume and anthropomorphic measurements: normal values," *Plastic & Reconstructive Surgery*, vol. 78, pp. 331-5, 1986.

[48] T. P. Brown, C. Ringrose, R. E. Hyland, A. A. Cole, and T. M. Brotherston, "A method of assessing female breast morphometry and its clinical application [comment]," *British Journal of Plastic Surgery*, vol. 52, pp. 355-9, 1999.

[49] M. Westreich, "Anthropomorphic breast measurement: protocol and results in 50 women with aesthetically perfect breasts and clinical application," *Plastic & Reconstructive Surgery*, vol. 100, pp. 468-79, 1997.

[50] B. Fowble, D. A. Fein, A. L. Hanlon, B. L. Eisenberg, J. P. Hoffman, E. R. Sigurdson, M. B. Daly, and L. J. Goldstein, "The impact of tamoxifen on breast recurrence, cosmesis, complications, and survival in estrogen receptor-positive early-stage breast cancer.," *International Journal of Radiation Oncology, Biology, Physics*, vol. 35, pp. 669-77, 1996.

[51] R. D. Pezner, M. P. Patterson, L. R. Hill, N. Vora, K. R. Desai, J. O. Archambeau, and J. A. Lipsett, "Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer," *International Journal of Radiation Oncology, Biology, Physics*, vol. 11, pp. 575-8, 1985.

[52] E. Van Limbergen, E. van der Schueren, and K. Van Tongelen, "Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system," *Radiotherapy & Oncology*, vol. 16, pp. 159-67, 1989.

[53] E. Van Limbergen, A. Rijnders, E. van der Schueren, T. Lerut, and R. Christiaens, "Cosmetic evaluation of breast conserving treatment for mammary cancer. 2. A quantitative analysis of the influence of radiation dose, fractionation schedules and surgical treatment techniques on cosmetic results," *Radiotherapy & Oncology*, vol. 16, pp. 253-67, 1989.

[54] D. B. Sheffer, T. E. Price, C. W. Loughry, B. L. Bolyard, W. M. Morek, and R. S. Varga, "Validity and reliability of biostereometric measurement of the human female breast," *Annals of Biomedical Engineering*, vol. 14, pp. 1-14, 1986.

[55] J. C. Boot, A. J. Naftel, and A. R. Ramli, "Bodymap: an image processing system for the measurement of body surface profiles encountered in skin expansion surgery," *International Journal of Bio-Medical Computing*, vol. 31, pp. 189-204, 1992.

[56] C. M. Malata, S. A. McIntosh, and A. D. Purushotham, "Immediate breast reconstruction after mastectomy for cancer," *British Journal of Surgery*, vol. 87, pp. 1455-72, 2000.

[57] G. M. Galdino, M. Nahabedian, M. Chiaramonte, J. Z. Geng, S. Klatsky, and P. Manson, "Clinical applications of three-dimensional photography in breast surgery," *Plastic & Reconstructive Surgery*, vol. 110, pp. 58-70, 2002.

[58] C. W. Loughry, D. B. Sheffer, T. E. Price, R. L. Einsporn, R. G. Bartfai, W. M. Morek, and N. M. Meli, "Breast volume measurement of 598 women using biostereometric analysis," *Annals of Plastic Surgery*, vol. 22, pp. 380-5, 1989.

[59] M. S. Kim, J. C. Sbalchiero, G. P. Reece, M. J. Miller, E. K. Beahm, and M. K. Markey, "Assessment of Breast Aesthetics," *Plastic & Reconstructive Surgery*, in press 2007.

[60] J. Bostwick, 3rd, "Reconstruction after mastectomy," *Surgical Clinics of North America*, vol. 70, pp. 1125-40, 1990.

[61] S. S. Bushnell, "Breast reconstruction after mastectomy," *Hematology - Oncology Clinics of North America*, vol. 3, pp. 709-25, 1989.

[62] H. S. Cody, 3rd, "Current surgical management of breast cancer," *Current Opinion in Obstetrics & Gynecology.*, vol. 14, pp. 45-52, 2002.

[63] R. DeBono, A. Thompson, and J. H. Stevenson, "Immediate versus delayed free TRAM breast reconstruction: an analysis of perioperative factors and complications.," *British Journal of Plastic Surgery*, vol. 55, pp. 111-6, 2002.

[64] E. A. Krueger, E. G. Wilkins, M. Strawderman, P. Cederna, S. Goldfarb, F. A. Vicini, and L. J. Pierce, "Complications and patient satisfaction following expander/implant breast reconstruction with and without radiotherapy," *International Journal of Radiation Oncology, Biology, Physics*, vol. 49, pp. 713-21, 2001.

[65] S. S. Kroll, *The well-informed patient's guide to breast reconstruction*. Houston: The University of Texas M. D. Anderson Cancer Center, 2002.

[66] S. S. Kroll, "Why autologous tissue?," *Clinics in Plastic Surgery*, vol. 25, pp. 135-43, 1998.

[67] S. L. Moran and J. M. Serletti, "Outcome comparison between free and pedicled TRAM flap breast reconstruction in the obese patient.," *Plastic & Reconstructive Surgery*, vol. 108, pp. 1954-60, 2001.

[68] J. M. Serletti and S. L. Moran, "Free versus the pedicled TRAM flap: a cost comparison and outcome analysis.," *Plastic & Reconstructive Surgery*, vol. 100, pp. 1425-7, 1997.

[69] K. C. Shestak, "Breast reconstruction with a pedicled TRAM flap," *Clinics in Plastic Surgery*, vol. 25, pp. 167-82, 1998.

[70] D. F. Veiga, M. SabinoNeto, L. M. Ferreira, E. B. Garcia, J. VeigaFilho, N. F. Novo, and J. L. Rocha, "Quality of life outcomes after pedicled TRAM flap delayed breast reconstruction," *British Journal of Plastic Surgery*, vol. 57, pp. 252-257, 2004.

[71] K. J. Walgenbach and K. C. Shestak, "Pedicled TRAM Breast Reconstruction," *Breast Disease*, vol. 16, pp. 73–77, 2002.

[72] J. D. Namnoum, "Pedicle Versus Free TRAM For Breast Reconstruction," *Breast Disease*, vol. 16, pp. 79–83, 2002.

[73] B. J. Baldwin, M. A. Schusterman, M. J. Miller, S. S. Kroll, and B. G. Wang, "Bilateral breast reconstruction: conventional versus free TRAM," *Plastic & Reconstructive Surgery*, vol. 93, pp. 1410-6; discussion 1417, 1994.

[74] M. A. Schusterman, S. S. Kroll, M. J. Miller, G. P. Reece, B. J. Baldwin, G. L. Robb, C. S. Altmyer, F. C. Ames, S. E. Singletary, and M. I. Ross, "The free transverse rectus abdominis musculocutaneous flap for breast reconstruction: one center's experience with 211 consecutive cases," *Annals of Plastic Surgery*, vol. 32, pp. 234-41; discussion 241-2, 1994.

[75] M. A. Schusterman, "The free TRAM flap," *Clinics in Plastic Surgery*, vol. 25, pp. 191-5, 1998.

[76] R. Allen, "DIEP versus TRAM for Breast Reconstruction," *Plastic and Reconstructive Surgery*, vol. 111, pp. 2478, 2003.

[77] G. R. Evans and S. S. Kroll, "Choice of technique for reconstruction," *Clin. Plast. Surg.*, vol. 25, pp. 311-6, 1998.

[78] M. Morrow, E. A. Strom, L. W. Bassett, D. D. Dershaw, B. Fowble, A. Giuliano, J. R. Harris, M. O, F., S. J. Schmidt, S. E. Singletary, and D. P. Winchester, "Standard for breast conservation therapy in the management of invasive breast carcinoma," *CA Cancer J. Clin.*, vol. 52, pp. 277-300, 2002.

[79] L. M. Apantaku, "Breast conserving surgery for breast cancer," *Am. Fam. Phy.*, vol. 66, pp. 2271-2278, 2002.

[80] R. M. Goldwyn, *Plastic and reconstructive surgery of the breast*. Boston: Little Brown, 1976.

[81] S. K. Al-Ghazal, L. Fallowfield, and R. W. Blamey, "Does cosmetic outcome from treatment of primary breast cancer influence psychosocial morbidity?," *Eu. J. Surg. Oncol.*, vol. 25, pp. 571-3, 1999.

[82] L. Kovacs, A. Zimmermann, N. A. Papadopulos, and E. Biemer, "Re: Factors determining shape and symmetry in immediate breast reconstruction.," *Ann. Plast. Surg.*, vol. 53, pp. 192-194, 2004.

[83] D. B. Sarwer, "Differences in breast shape preferences between plastic surgeons and patients seeking breast augmentation [Discussion]," *Plast. Reconstr. Surg.*, 2003.

[84] H. C. Hsia and J. G. Thomson, "Differences in breast shape preferences between plastic surgeons and patients seeking breast augmentation," *Plast. Reconstr. Surg.*, vol. 112, pp. 312-320, 2003.

[85] D. B. Sarwer, T. F. Cash, L. Magee, E. F. Williams, J. K. Thompson, M. Roehrig, S. Tantleff-Dunn, A. K. Agliata, D. E. Wilfley, A. D. Amidon, D. A. Anderson, and M. Romanofski, "Female college students and cosmetic surgery: an investigation of experiences, attitudes, and body image," *Plast. Reconstr. Surg.*, vol. 115, pp. 931-938, 2005.

[86] T. von Soet, I. L. Kvalem, K. C. Skolleborg, and H. E. Roald, "Psychosocial factors predicting the motivation to undergo cosmetic surgery," *Plast. Reconstr. Surg.*, vol. 117, pp. 51-62, 2006.

[87] D. B. Sarwer, "Psychosocial factors predicting the motivation to undergo cosmetic surgery [Discussion]," *Plast. Reconstr. Surg.*, vol. 117, pp. 63-64, 2006.

[88] W. N. Andrade and J. L. Semple, "Patient self-assessment of the cosmetic results of breast reconstruction. ," *Plastic & Reconstructive Surgery*, vol. 117, pp. 44-47, 2006.

[89] D. B. Sarwer, "Patient self-assessment of the cosmetic results of breast reconstruction [Discussion] " *Plast. Reconstr. Surg.*, vol. 117, pp. 48-49, 2006.

[90] D. B. Sarwer, T. Pruzinsky, T. F. Cash, R. M. Goldwyn, J. A. Persing, and L. A. Whitaker, *Psychological aspects of reconstructive and cosmetic plastic surgery: clinical, empirical, and ethical perspectives/editors.* Philadelphia: Lippincott Williams & Wilkins, 2005.

[91] A. Edsander-Nord, Y. Brandberg, and M. Wickman, "Quality of life, patients' satisfaction, and aesthetic outcome after pedicled or free TRAM flap breast surgery," *Plast. Reconstr. Surg.*, vol. 107, pp. 1142, 2001.

[92] S. Asko-Seljavaara, "Quality of life, patients' satisfaction, and aesthetic outcome after pedicled or free TRAM flap breast surgery [discussion]," *Plast, Reconstr. Surg.*, vol. 107, pp. 1154-5, 2000.

[93] R. D. Pezner, J. A. Lipsett, N. L. Vora, and K. R. Desai, "Limited usefulness of observer-based cosmesis scales employed to evaluate patients treated

conservatively for breast cancer," *Int. J. Rad. Onc. Bio. Phy.*, vol. 11, pp. 1117-9, 1985.

[94] J. C. Lowery, E. G. Wilkins, W. M. Kuzon, and J. A. Davis, "Evaluations of aesthetic results in breast reconstruction: an analysis of reliability," *Ann. Plast. Surg.*, vol. 36, pp. 601-6; discussion 607, 1996.

[95] K. C. Sneeuw, N. K. Aaronson, J. R. Yarnold, M. Broderick, J. Regan, G. Ross, and A. Goddard, "Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. Comparison of patients' ratings, observers' ratings, and objective assessments," *Radiother. Oncol.*, vol. 25, pp. 153-9, 1992.

[96] J. Boyages, B. Barraclough, J. Middledorp, D. Gorman, and A. O. Langlands, "Early breast cancer: cosmetic and functional results after treatment by conservative techniques.," *Aus. New Zealand J. Surg.*, vol. 58, pp. 111-21, 1988.

[97] D. Clarke, A. Martinez, and R. S. Cox, "Analysis of cosmetic results and complications in patients with stage I and II breast cancer treated by biopsy and irradiation," *Int. J. Rad. Oncol. Biol. Phy.*, vol. 9, pp. 1807-13, 1983.

[98] H. Barterlink, F. van Dam, and J. van Dogen, "Psychological effects of breast conserving therapy in comparison with radical mastectomy," *Int. J. Rad. Oncol. Biol. Phy.*, vol. 11, pp. 381-385, 1985.

[99] M. Cohen, B. Evanoff, L. T. George, and K. E. Brandt, "A subjective rating scale for evaluating the appearance outcome of autologous breast reconstruction.," *Plastic and Reconstructive Surgery*, vol. 116, pp. 440-9, 2005.

[100] M. Cohen, B. Evanoff, L. T. George, and K. E. Brandt, "A subjective rating scale for evaluating the appearance outcome of autologous breast reconstruction.," *Plastic & Reconstructive Surgery*, vol. 116, pp. 440-9, 2005.

[101] J. Penn, "Breast Reduction," *Br. J. Plast. Surg.*, vol. 7, pp. 357, 1955.

[102] D. J. Smith, Jr., W. E. Palin, Jr., V. L. Katch, and J. E. Bennett, "Breast volume and anthropomorphic measurements: normal values," *Plast. Reconstr. Surg.*, vol. 78, pp. 331-5, 1986.

[103] Penn, "Breast Reduction," *British Journal of Plastic Surgery*, vol. 7, pp. 357, 1955.

[104] D. J. Hauben, N. Adler, R. Silfen, and D. Regev, "Breast-Areola-Nipple Proportion," *Annals of Plastic Surgery*, vol. 50, pp. 510-513, 2003.

[105] D. J. Smith, Jr., W. E. Palin, Jr., V. Katch, and J. E. Bennett, "Surgical treatment of congenital breast asymmetry," *Ann. Plast Surg.*, vol. 17, pp. 92-101, 1986.

[106] L. I. Tsouskas and I. S. Fentiman, "Breast compliance: a new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer," *Breast Cancer Res. Treat.*, vol. 15, pp. 185-90, 1990.

[107] B. Stark and N. Olivari, "Breast asymmetry: an objective analysis of postoperative results," *Eu. J. Plast. Surg.*, vol. 14, pp. 173-176, 1991.

[108] R. Kuzbari, M. Deutinger, B. P. Todoroff, B. Schneider, and G. Freilinger, "Surgical treatment of developmental asymmetry of the breast. Long term results," *Scand. J. Plast. Reconstr. Surg. Hand Surg.*, vol. 27, pp. 203-7, 1993.

[109] D. J. Hauben, N. Adler, R. Silfen, and D. Regev, "Breast-areola-nipple proportion," *Ann. Plast. Surg.*, vol. 50, pp. 510-513, 2003.

[110] S. K. Al-Ghazal, R. W. Blamey, J. Stewart, and A. A. Morgan, "The cosmetic outcome in early breast cancer treated with breast conservation," *Eu. J. Surg. Oncol.*, vol. 25, pp. 566-70, 1999.

[111] C. M. Malata, S. A. McIntosh, and A. D. Purushotham, "Immediate breast reconstruction after mastectomy for cancer," *Br. J. Surg.*, vol. 87, pp. 1455-72, 2000.

[112] C. Eadie, A. Herd, and S. Stallard, "An investigation into digital imaging in assessing cosmetic outcome after breast surgery," *J. Audiovisual Media Med.*, vol. 23, pp. 12-6, 2000.

[113] M. S. Kim, G. P. Reece, M. J. Miller, E. Beahm, E. N. Atkinson, and M. K. Markey, "Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," *Computers in Biology and Medicine*, vol. 37, pp. 49-59, 2007.

[114] E. Van Limbergen, A. Rijnders, E. van der Schueren, T. Lerut, and R. Christiaens, "Cosmetic evaluation of breast conserving treatment for mammary cancer. 2. A quantitative analysis of the influence of radiation dose, fractionation schedules and surgical treatment techniques on cosmetic results," *Radiother. Oncol.*, vol. 16, pp. 253-67, 1989.

[115] E. Van Limbergen, E. van der Schueren, and K. Van Tongelen, "Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system," *Radiother. Oncol.*, vol. 16, pp. 159-67, 1989.

[116] C. Vrieling, L. Collette, E. Bartelink, J. H. Borger, S. J. Brenninkmeyer, J. C. Horiot, M. Pierart, P. M. Poortmans, H. Struikmans, E. Van der Schueren, J. A. Van Dongen, E. Van Limbergen, and H. Bartelink, "Validation of the methods of cosmetic assessment after breast-conserving therapy in the EORTC "boost versus no boost" trial. EORTC radiotherapy and breast cancer cooperative groups. european organization for research and treatment of cancer," *Int. J. Rad. Onc. Bio. Phy.*, vol. 45, pp. 667-76, 1999.

[117] R. D. Pezner, M. P. Patterson, L. R. Hill, N. Vora, K. R. Desai, J. O. Archambeau, and J. A. Lipsett, "Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer," *Int. J. Rad. Oncol. Bio. Phy.*, vol. 11, pp. 575-8, 1985.

[118] V. Sacchini, A. Luini, S. Tana, L. Lozza, V. Galimberti, M. Merson, R. Agresti, P. Veronesi, and M. Greco, "Quantitative and qualitative cosmetic evaluation after conservative treatment for breast cancer," *Eu. J. Cancer*, vol. 27, pp. 1395-400, 1991.

[119] P. Regnault, "Breast ptosis. definition and treatment.," *Clin. Plast Surg.*, vol. 3, pp. 193-203, 1976.

[120] M. S. Kim, G. P. Reece, M. J. Miller, E. Beahm, E. N. Atkinson, and M. K. Markey, "Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," *Comp. Biol. Med.*, vol. in press, 2005.

[121] R. Yavuzer, S. Smirnes, and I. T. Jackson, "Guidelines for standard photography in plastic surgery," *Plastic & Reconstructive Surgery*, vol. 46, pp. 293-300, 2001.

[122] H. A. Zarem, "Standards of Photography," *Plastic & Reconstructive Surgery*, vol. 74, 1984.

[123] "Photographic Standards in Plastic Surgery," Plastic Surgery Educaton Foundation Education Tecnologies Committee 2000.

[124] B. E. DiBernardo, R. L. Adams, J. Krause, M. A. Fiorillo, and G. Gheradini, "Photographic Standards in Plastic Surgery," *Plastic & Reconstructive Surgery*, vol. 102, pp. 559-568, 1998.

[125] R. E. Ellenbogen, S. J. Jankauskas, and F. J. and Collini, "Achieving standardized potographs in aesthetic surgery," *Plastic & Reconstructive Surgery*, vol. 86, pp. 955-958, 1989.

[126] C. P. Honrado and W. F. J. Larrabee, "Update in three-dimensional imaging in facial plastic surgery.," *Cur. Opi. Otolaryngol Head Neck Surg.*, vol. 12, pp. 327-31, 2004.

[127] K. F. O'Grady and O. M. Antonyshyn, "Facial Asymmetry: Three-Dimensional Analysis Using Laser Surface Scanning," *Plastic & Reconstructive Surgery*, vol. 104, pp. 928-937, 1999.

[128] A. C. Da Silveira, J. L. J. Daw, B. Kusnoto, C. Evans, and M. Cohen, "Craniofacial applications of three-dimensional laser surface scanning.," *J. Craniofacial Surg.*, vol. 14, pp. 449-56, 2003.

[129] E. Nkenke, A. Langer, X. Laboureux, M. Benz, T. Maier, M. Kramer, G. Hausler, P. Kessler, J. Wiltfang, and F. W. Neukam, "Validation of in vivo assessment of facial soft-tissue volume changes and clinical application in midfacial distraction: a technical report.," *Plastic & Reconstructive Surgery*, vol. 112, pp. 367-80, 2003.

[130] V. F. Ferrario, C. Sforza, C. Dellavia, G. M. Tartaglia, A. Colombo, and A. Caru, "A quantitative three-dimensional assessment of soft tissue facial asymmetry of cleft lip and palate adult patients.," *J. Craniofacial Surg.*, vol. 14, pp. 739-46, 2003.

[131] L. Samson, "Three-Dimensional Photography and Its Application to Facial Plastic Surgery," *Arch. facial plast. surg.*, vol. 6, pp. 410-414, 2004.

[132] J. C. Boot, A. J. Naftel, and A. R. Ramli, "Bodymap: an image processing system for the measurement of body surface profiles encountered in skin expansion surgery," *Int. J. Bio-Med. Comp.*, vol. 31, pp. 189-204, 1992.

[133] C. M. Malata, J. C. Boot, E. T. Bradbury, A. R. Ramli, and D. T. Sharpe, "Congenital breast asymmetry: subjective and objective assessment," *Br. J. Plast. Surg.*, vol. 47, pp. 95-102, 1994.

[134] C. W. Loughry, D. B. Sheffer, T. E. Price, Jr., M. J. Lackney, R. G. Bartfai, and W. M. Morek, "Breast volume measurement of 248 women using biostereometric analysis," *Plast. Reconstr. Surg.*, vol. 80, pp. 553-8, 1987.

[135] A. Losken, H. Seify, D. D. Denson, A. A. J. Paredes, and G. W. Carlson, "Validating three-dimensional imaging of the breast.," *Ann. Plast. Surg.*, vol. 54, pp. 471-6, 2005.

[136] A. J. Grossman and L. A. and Rounder, "A simple means for accurate breast volume determination," *Plastic & Reconstructive surgery*, vol. 66, 1980.

[137] D. B. Sheffer, T. E. Price, C. W. Loughry, B. L. Bolyard, W. M. Morek, and R. S. Varga, "Validity and reliability of biostereometric measurement of the human female breast," *Ann. Biomed. Eng.*, vol. 14, pp. 1-14, 1986.

[138] C. W. Loughry, D. B. Sheffer, T. E. Price, Jr., R. L. Einsporn, R. G. Bartfai, W. M. Morek, and N. M. Meil, "Breast volume measurement of 598 women using biostereometric analysis," *Ann. Plast Surg.*, vol. 22, pp. 380-385, 1989.

[139] W. E. J. Palin, J. A. von Fraunhofer, and D. J. J. Smith, "Measurement of breast volume: comparison of techniques.," *Plastic & Reconstructive Surgery*, vol. 77, pp. 253-5, 1986.

[140] A. Edsander-Nord, M. Wickman, and G. Jurell, "Measurement of breast volume with thermoplastic casts.," *Scan. J. Plast. Reconstr. Surg. Hand Surg.*, vol. 30, pp. 129-32, 1996.

[141] E. Tezel and A. Numanoglu, "Practical do-it-yourself device for accurate volume measurement of breast.," *Plastic & Reconstructive Surgery*, vol. 105, pp. 1019-23, 2000.

[142] C. L. Kalbhen, J. J. McGill, P. M. Fendley, K. W. Corrigan, and J. Angelats, "Mammographic determination of breast volume: comparing different methods.," *Am. J. Roent.*, vol. 173, pp. 1643-9, 1999.

[143] M. Mineyev, D. Kramer, L. Kaufman, J. Carlson, and S. Frankel, "Measurement of breast implant volume with magnetic resonance imaging.," *Ann. Plast. Surg.*, vol. 34, pp. 348-51, 1995.

[144] L. E. Edstrom, M. C. Robson, and J. K. Wright, "A method for the evaluation of minor degrees of breast asymmetry," *Plastic & Reconstructive Surgery*, vol. 60, pp. 812-4, 1977.

[145] C. Rigotti, G. Ferrigno, A. Aliverti, and A. Pedotti, "Surface scanning: an application to mammary surgery," *J. Biomed. Optics*, vol. 3, pp. 161-170, 1998.

[146] G. T. Passalis, M. M. Theoharis, and I. A. Kakadiaris, "Non-invasive automatic breast volume estimation for post-mastectomy breast reconstructive surgery," presented at Proc. 25th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc., 2003.

[147] R. A. Jacobs, "Three-dimensional photography.," *Plastic & Reconstructive Surgery*, vol. 107, pp. 276-7, 2001.

[148] A. Losken, H. Seify, D. D. Denson, A. A. J. Paredes, and G. W. Carlson, "Validating three-dimensional imaging of the breast.," *Annals of Plastic Surgery*, vol. 54, pp. 471-6, 2005.

[149] L. Kovacs, A. Zimmermann, N. A. Papadopulos, and E. Biemer, "Re: Factors determining shape and symmetry in immediate breast reconstruction.," *Annals of Plastic Surgery*, vol. 53, pp. 192-194, 2004.

[150] M. Y. Nahabedian, "Invited discussion : validating three-dimensional imaging of the breast," *Ann. Plast. Surg.*, vol. 54, pp. 477-478, 2005.

[151] A. Bayat, D. A. McGrouther, and M. W. J. Ferguson, "Skin scarring," *BMJ (Clinical Research Ed.)*, vol. 326, pp. 88-92, 2003.

[152] B. S. Atiyeh, J. Ioannovich, C. A. Al-Amm, K. A. El-Musa, and R. Dham, "Improving scar quality: a prospective clinical study," *Aesthetic Plastic Surgery*, vol. 26, pp. 470-476, 2002.

[153] Y. Godwin, "A comparison of the patient and surgeon opinion on the long-term aesthetic outcome of reduction mammaplasty," *British Journal of Plastic Surgery*, vol. 51, pp. 444-449, 1998.

[154] P. P. M. van Zuijlen, A. P. Angeles, R. W. Kreis, K. E. Bos, and E. Middelkoop, "Scar assessment tools: implications for current research," *Plastic And Reconstructive Surgery*, vol. 109, pp. 1108-1122, 2002.

[155] D. F. Veiga, M. S. Neto, Y. Juliano, L. M. Ferreira, and J. L. B. S. Rocha, "Evaluations of the aesthetic results and patient satisfaction with the late pedicled TRAM flap breast reconstruction," *Annals of Plastic Surgery*, vol. 48, pp. 515-520, 2002.

[156] A. L. van de Kar, L. U. Corion, S. M. J., L. J. Draaijers, C. M. van der Horst, and P. P. van Zuijlen, "Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale," *Plastic & Reconstructive Surgery*, vol. 116, pp. 514-22, 2005.

[157] P. T. Truong, F. Abnousi, C. M. Yong, A. Hayashi, J. A. Runkel, T. Phillips, and I. A. Olivotto, "Standardized assessment of breast cancer surgical scars integrating the Vancouver Scar Scale, Short-Form McGill Pain Questionnaire, and patients' perspectives.," *Plastic And Reconstructive Surgery*, vol. 116, pp. 1291-1299, 2005.

[158] T. Sullivan, J. Smith, J. Kermode, E. McIver, and D. J. Courtemanche, "Rating the burn scar.," *Journal of Burn Care and Rehabilitation*, vol. 11, pp. 256-60, 1990.

[159] L. J. Draaijers, F. R. H. Tempelman, Y. A. M. Botman, R. W. Kreis, E. Middelkoop, and P. P. M. van Zuijlen, "Colour evaluation in scars: tristimulus colorimeter, narrow-band simple reflectance meter or subjective evaluation?," *Burns: Journal Of The International Society For Burn Injuries*, vol. 30, pp. 103-107, 2004.

[160] M. Setaro and A. Sparavigna, "Quantification of erythema using digital camera and computer-based colour image analysis: a multicentre study," *Skin research and technology*, pp. 84-88, 2002.

[161] S. G. Coelho, S. A. Miller, B. Z. Zmudzka, and J. Z. Beer, "Quantification of UV-induced erythema and pigmentation using computer-assisted digital image evaluation," *Photochemistry and Photobiology*, vol. 82, pp. 651-655, 2006.

[162] P. P. M. van Zuijlen, A. P. Angeles, M. H. Suijker, R. W. Kreis, and E. Middelkoop, "Reliability and accuracy of techniques for surface area measurements of wounds and scars," *The International Journal Of Lower Extremity Wounds*, vol. 3, pp. 7-11, 2004.

[163] T. Hurtut, F. Cheriet, J. Joncas, and J. Dansereau, "Enhancement and Segmentation of Scar Color Images after a Scoliosis Surgery," presented at Digital Image Computing: Techniques and Applications, Sydney, 2003.

[164] L. J. Draaijers, F. R. H. Tempelman, Y. A. M. Botman, W. E. Tuinebreijer, E. Middelkoop, R. W. Kreis, and P. P. M. van Zuijlen, "The Patient and Observer Scar Assessment Scale: A Reliable and Feasible Tool for Scar Evaluation," *Plastic & Reconstructive Surgery June*, vol. 113, pp. 1960-1965, 2004.

[165] M. J. Baryza and G. A. Baryza, "The Vancouver Scar Scale: an administration tool and its interrater reliability.," *Journal of Burn Care & Rehabilitation.*, vol. 16, pp. 535-8, 1995.

[166] E. Beausang, H. Floyd, K. W. Dunn, C. I. Orton, and M. W. J. Ferguson, "A New Quantitative Scale for Clinical Scar Assessment," *Plastic & Reconstructive Surgery*, vol. 102, pp. 1954-1961, 1998.

[167] L. Andreassi and L. Flori, "Practical applications of cutaneous colorimetry," *Clinics In Dermatology*, vol. 13, pp. 369-373, 1995.

[168] H. Takiwaki, "Measurement of skin color: practical application and theoretical considerations," *The Journal of Medical Investigation*, vol. 44, pp. 121-126, 1998.

[169] S. G. Coelho, S. A. Miller, B. Z. Zmudzka, and J. Z. Beer, "Quantification of UV-induced erythema and pigmentation using computer-assisted digital image evaluation," *Photochem Photobiol*, vol. 82, pp. 651-655, 2006.

[170] H. Takiwaki, Y. Kanno, Y. Miyaoka, and S. Arase, "Computer simulation of skin color based on a multilayered skin model," *skin research And Technology*, vol. 3, pp. 36-41, 1997.

[171] S. Wan, R. R. Anderson, and J. A. Parrish, "Analytical modeling for the optical properties of the skin with in vitro and in vivo applications," *Photochem Photobiol*, vol. 34, pp. 493-9, 1981.

[172] S. Wellek, *Testing statistical hypothesis of equivalence*. Boca Raton, FL: CRC Press LLC, 2003.

[173] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological Methods*, vol. 1, pp. 30-46, 1996.

[174] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, pp. 307-310, 1986.

[175] J. M. Bland and D. G. Altman, "Applying the right statistics: analyses of measurement studies," *Ultrasound in obstetrics & gynecology*, vol. 22, pp. 85-93, 2003.

[176] S. L. Moran and J. M. Serletti, "Outcome comparison between free and pedicled TRAM flap breast reconstruction in the obese patient," *Plastic & Reconstructive Surgery*, vol. 108, pp. 1954-1960, 2001.

[177] J. M. Serletti and S. L. Moran, "Free versus the pedicled TRAM flap: a cost comparison and outcome analysis.," *Plastic & Reconstructive Surgery*, vol. 100, pp. 1425-1427, 1997.

[178] K. C. Shestak, "Breast reconstruction with a pedicled TRAM flap," *Clinics in Plastic Surgery*, vol. 25, pp. 167-182, 1998.

[179] T. B. Fitzpatrick, "Soleil et peau," *Journal of Medical Esthetics*, vol. 2, pp. 33-34, 1975.

[180] P. Regnault, "Breast ptosis: definition and treatment.," *Clinics in Plastic Surgery*, vol. 3, pp. 193-203, 1976.

[181] J. Serup and T. Agner, "Colorimetric quantification of erythema--a comparison of two colorimeters (Lange Micro Color and Minolta Chroma Meter CR-200) with a clinical scoring scheme and laser-Doppler flowmetry," *Clinical And Experimental Dermatology*, vol. 15, pp. 267-272, 1990.

[182] E. Van den Kerckhove, F. Staes, M. Flour, K. Stappaerts, and W. Boeckx, "Reproducibility of repeated measurements on healthy skin with Minolta Chromameter CR-300," *Skin Research and Technology*, vol. 7, pp. 56-59, 2001.

[183] H. Takiwaki, L. Overgaard, and J. Serup, "Comparison of narrow-band reflectance spectrophotometric and tristimulus colorimetric measurements of skin color. Twenty-three anatomical sites evaluated by the Dermaspectrometer and the Chroma Meter CR-200," *Skin Pharmacology*, vol. 7, pp. 217-225, 1994.

[184] C. W. Li-Tsang, J. C. Lau, and S. K. Liu, "Validation of an objective scar pigmentation measurement by using a spectrocolorimeter," *Burns*, vol. 29, pp. 779-84, 2003.

[185] L. Minolta Co., "Chroma Meter CR-300/CR-310/CR-321/CR-331/CR-331C Instruction Manual."

[186] R. R. Luiz and M. Szklo, "More than one statistical strategy to assess agreement of quantitative measurements may usefully be reported," *Journal of clinical epidemiology*, vol. 58, pp. 215-216, 2005.

[187] B. Rosner, *Fundamentals of Biostatistics*, 6th ed: Duxbury Press, 2005.

[188] M. D. Shriver and E. J. Parra, "Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry," *Am J Phys Anthropol*, vol. 112, pp. 17-27, 2000.

[189] M. S. Kim, G. P. Reece, M. J. Miller, E. Beahm, E. N. Atkinson, and M. K. Markey, "Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," *Computers in Biology and Medicine*, vol. 37, pp. 49-59, 2007.

[190] S. Wellek, *Testing statistical hypothesis of equivalence*. Boca Raton, FL: CRC Press LLC, 2003.

[191] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet.*, vol. 1, pp. 307-10, 1986.

[192] J. M. Bland and D. G. Altman, "Applying the right statistics: analyses of measurement studies," *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology.*, vol. 22, pp. 85-93, 2003.

[193] J. C. Nunnally and I. H. Berstein, *Psychometric Theory*, third ed. New York: McGraw-Hill Inc., 1994.

[194] R. F. DeVellis, *Scale Development*, second ed. Thousand Oaks CA: Sage Publications Inc., 2003.

[195] L. Crocker and J. Algina, *Introduction to Classical and Modern Test Theory*. Belmont CA: Wadsworth Group/Thomson Learning, 1986.

[196] L. Itti, "Computational Modelling of Visual Attention," *Neuroscience*, vol. 2, pp. 194-203, 2001.

[197] C. Mello-Thoms and L. Hardesty, "How Does Lesion Conspicuity Affect Visual Search Strategy in Mammogram Reading?," presented at SPIE Medical Imaging, San Diego, 2005.

[198] E. A. Krupinski, "Visual scanning patterns of radiologists searching mammograms," *Academic Radiology*, vol. 3, pp. 137-44, 1996.

[199] E. A. Krupinski, "Visual search of mammographic images: influence of lesion subtlety," *Academic radiology.*, vol. 12, pp. 965-9, 2005.

[200] C. Mello-Thoms, S. Dunn, C. F. Nodine, H. L. Kundel, and S. P. Weinstein, "The Perception of Breast Cancer: What Differentiates Missed from Reported Cancers in Mammography?1 " *Academic Radiology*, vol. 9, pp. 1004-1012, 2002.

[201] A. M. Herbert, J. B. Pelz, L. Calderwood, M. Cook, M. Curtis, C. DeAngelis, and B. Garrison, "Searching for symmetry: Eye movements during a difficult symmetry detection task," *Journal of Vision*, vol. 6, pp. 24a, 2006.

[202] R. L. Mappus IV, R. W. Ferguson, K. Czechowski, and P. M. Corballis, "Spotting differences: How qualitative asymmetries influence visual search," presented at Cognitive Science Conference, Stresa, Italy, 2005.

Vita

Min Soon Kim was born in Daejeon, South Korea on August 20th, 1974 as the son of Sunhong Kim and Wonja Kim. He entered Chungnam National University, Daejeon, South Korea in 1993. He completed his B.S. degree in the School of Mechanical Engineering at Chungnam National University in August 2000. He began his M.S. degree in the Department of Mechanical Engineering at the University of Texas at Austin in September 2001 and transferred to The University of Texas Department of Biomedical Engineering in October 2002 and received his M.S. degree in December 2005. He joined the Biomedical Informatics Lab at The University of Texas at Austin in October 2002 and has developed algorithms for objective assessment of aesthetic outcomes of breast cancer treatment: quantifying aesthetic factors after breast reconstruction under Dr. Mia K. Markey's supervision.

Permanent address: Samsung Apt 3-1405, Oryu-dong 175-1, Jung-gu, Daejeon, South Korea 301-758

This dissertation was typed by Min Soon Kim.